DOI: 10.1111/phpe.12194

ORIGINAL ARTICLE

PHILOSOPHICAL PERSPECTIVES WILEY

Are there really any dual-character concepts?

Jonathan Phillips | **David Plunkett**

Dartmouth College

Correspondence
Jonathan Phillips, Dartmouth College.
Email: Jonathan.S.Phillips@dartmouth.edu

Abstract

There has been growing excitement in recent years about "dual-character" concepts. Philosophers have argued that such concepts can help us make progress on a range of philosophical issues, from aesthetics to law to metaphysics. Dual-character concepts are thought to have a distinctive internal structure, which relates a set of descriptive features to an abstract value, and which allows people to use either the descriptive features or the abstract value for determining the extension of the concept. Here, we skeptically investigate the central argument in favor of their existence. Across three new empirical studies, we systematically demonstrate that the linguistic patterns that dual-character concepts were originally posited to explain are likely better explained by much more general features of language use.

1 | INTRODUCTION

In a 2013 paper entitled "Dual-Character Concepts and the Normative Dimension of Representation", Joshua Knobe, Sandeep Prasada, and George Newman argued that we need to recognize a distinct class of concepts, which they dubbed "dual-character concepts".[1] According to them, dual-character concepts "involve two ways of characterizing their instances, and thus two ways of determining category membership."[2] Specifically, they argue that dual-character concepts involve one way of characterizing membership in terms of "a set of concrete features" and another in terms of an "abstract value".[3] On their view, "these two characterizations are intrinsically related, but they are nonetheless distinct, and they can sometimes yield opposing verdicts about whether a particular object counts as a category member or not."[4]

To illustrate their basic idea, consider their leading example of the concept SCIENTIST.[5] They argue that, on the one hand, people characterize someone as a scientist based on "concrete features" such as "conducting experiments", "analyzing data", "developing theories", and "writing papers".[6] On the other hand, they claim that someone having these features generally helps to

realize an abstract value – namely, "*the pursuit of empirical knowledge*"[7] – that can itself be directly used to classify people as scientists or not. The "dual-character" nature of SCIENTIST, and other concepts like it, is meant to arise from a distinctive internal structure where the descriptive criteria are generally ways of realizing an abstract value. In paradigm cases, such as Charles Darwin or Marie Curie, categorizing people using the concrete features and doing so using the abstract value both lead to the same verdict about whether someone is a scientist or not. In other cases, they pull apart. For example, someone might do all the tasks included in the "concrete features" but do them in a cynical way for money and fame, and not really care about the pursuit of empirical knowledge. Or, conversely, someone might be deeply invested in the pursuit of empirical knowledge, but not engage in the specific tasks glossed above, perhaps because she has yet to acquire the relevant skills and training needed to do so. Knobe, Prasada, and Newman claim that, in these cases where the criteria pull in different directions, it's not that the concept fails to apply, or is fundamentally "defective" in some sense.[8] Rather, people can (and do) still use the concept SCIENTIST to either classify the person as a scientist or not, but that classification will vary depending on which set of criteria they (at least implicitly) privilege.

Since Knobe, Prasada, and Newman first published their paper, they and others have further developed the claim that a number of important concepts have a "dual character". These include concepts central to long-standing philosophical debates, including PERSONAL IDENTITY, ART, LAW, certain social role concepts (involving gender, for example), and certain natural kind concepts.[9] Claims that these and other concepts are dual-character concepts are often meant to have substantial theoretical bite. For example, Sarah-Jane Leslie argues that positing the existence of dual-character concepts can help explain important patterns in our judgments involving "normative generics", and, in so doing, help us better understand the social/political dynamics of social identity (including gender identity).[10] In another vein, Sally Haslanger draws on the idea of dual-character concepts to explain key features of her views on philosophical methodology, regarding issues in "conceptual engineering" concerning the critical evaluation and improvement of concepts.[11] Or, to take a third example, Kevin Tobia, Newman, and Knobe argue that certain natural kind terms have a "dual-character structure". They argue that this helps explain our patterns of judgments about the chemical structure of water, connected to Hilary Putnam's "twin earth" thought experiment, and, in so doing, also helps us in our assessment of the philosophical theories based on it.[12]

Importantly, theorists working on dual-character concepts have not only argued that certain key concepts are "dual-character" ones, but many have also advanced new theories about what exactly dual-character concepts *are*. For example, a key debate is about what exactly the "internal structure" of dual-character concepts consists in: roughly, how the "concrete features" and the relevant normative dimension of a dual-character concept are related to each other. Above, we gave the original, canonical statement of what that internal structure consists in, which Prasada, Newman, and Knobe provided in their original paper. In more recent work, however, others (including Leslie, Guillermo Del Pinal, and Kevin Reuter) have argued for amendments to that structure.[13] For example, Leslie argues that, for dual-character concepts involving social kinds, the normative dimension of the concept is tied to (purported) ideals involved in the primary social role or function of that kind.[14] As our brief overview above illustrates, there is a substantial (and growing) literature both applying and refining the idea of dual-character concepts.

Despite all of this, we are not convinced that there are actually any such things as dual-character concepts. Our reason for skepticism comes from a deeper investigation of the core empirical patterns about language use that proponents of dual-character concepts have sought to

explain. In what follows, we argue that a clearer picture of the distinctive linguistic patterns that "dual-character" concepts were posited to explain suggests that the relevant linguistic patterns are actually not particularly distinctive of dual-character concepts after all. We build our case using three new studies we conducted, as well as by more carefully looking at the relevant kind of linguistic patterns Knobe, Prasada, and Newman originally used to motivate the existence of dual-character concepts. We argue that, rather than being explained by the existence of dual-character concepts, the overall patterns in linguistic behavior we observe might well be better explained by general features of how we use language in flexible ways across different contexts. Based on this, we don't see a strong reason to posit the existence dual-character concepts at work in our thought – or at least not given some relatively modest ideas about what it would take to be a distinct "type" of concept – and we might even have good reason to deny their existence.

2 | THE MAIN ARGUMENT FOR DUAL-CHARACTER CONCEPTS

In order to understand our argument in what follows, it will be helpful to have a more detailed account of the original argument that Knobe, Prasada, and Newman (and now others) use to support the existence of dual-character concepts. (Note that, as we just did here, we mostly use the term 'dual-character' non-quotationally, for ease of expression. However, as our argument will make clear, this choice should not be taken to involve a commitment on our part to the actual existence of dual-character concepts).

As we understand it, Knobe, Prasada, and Newman's main argument for dual-character concepts is that we should posit their existence in order to explain certain patterns in linguistic behavior. Put roughly, the core patterns here concern judgments of whether or not to apply a given term (e.g., 'scientist') to a given entity (e.g., a given person). Knobe, Prasada, and Newman posit a theory – namely, that certain concepts are dual-character ones, with a certain distinctive kind of internal structure – to account for the data. The idea – again, put roughly – is that certain terms (e.g., 'scientist') are tied to concepts (e.g., SCIENTIST) that have a "dual-character", whereas other terms are tied to concepts that do not. The argument for the existence of dual-character concepts is thus based on a certain familiar kind of "inference to best explanation". The idea is that positing that certain concepts are "dual-character concepts" is the best explanation for the relevant empirical data about linguistic behavior that they uncover.

One feature of this explanation that Knobe et al. don't stress but which we think is important is that they put things in terms of *concepts*. Knobe et al. don't make explicit what they take concepts to be, nor how they take them to be tied to words. But given that Knobe et al. take their findings to show us something important about cognition, and not just how we use language, we think it is charitable to read them as accepting a view of concepts (common to much of the contemporary philosophical literature) where they are understood to be things at the level of cognition (e.g., something like constituent components of thoughts). Moreover, given how smoothly Knobe et al. move between discussing words and concepts, we think it is charitable to take them as holding that there is a relatively tight connection between the two. For example: one such connection would be that the meanings of words can be explained in terms of the concepts they express.[15] For ease of exposition in what follows, we will start by following them in moving rather freely between talk of concepts and talk of words.

The core linguistic data that Knobe, Prasada, and Newman focus on involves test of how people respond to a series of assertions involving a given term (e.g., 'scientist') or a conventionalized combination of terms (e.g., 'rock music'). In these tests, participants first read a vignette

with a description of someone (or something) that was described as possessing some of the descriptive properties typically associated with a term but lacking some other evaluative or normative property also typically associated with the term. Some of the descriptions concern terms (such as 'scientist') that Knobe et al. take to be tied to dual-character concepts, while others concerned terms that (by hypothesis) are not tied to dual-character concepts.

To illustrate, consider the two descriptions below, one aimed at the term 'scientist', and one aimed at the term 'pharmacist' (which Knobe et al. take to be tied to a "control" concept, lacking a dual character):

> **'Scientist' description**: George is employed at Ameritech to run experimental studies and analyze the data. However, he actually has no interest at all in finding the correct answers to the questions he is studying. So although he goes through the motions, he does not actually care in any way about making a contribution to people's understanding of these issues.

> **'Pharmacist' description**: Laura has spent the last 10 years working at the local pharmacy. At work, she wears a white coat and fills medical prescriptions for her customers. She explains to customers how much medicine to take and when to do so. Furthermore, she warns patients about potentially dangerous interactions between drugs. Laura has no interest in medicine or helping people get well, but she likes the pay and benefits of her profession and wants to make sure she doesn't lose her job.

After reading each description, participants were asked to rate the truth of two assertions on a 7-point scale, as follows:

> **Member assertion**: There is a sense in which this person is a scientist [pharmacist].

> **Non-member assertion**: Ultimately when you think about what it really means to be a scientist [pharmacist], you would have to say that this person is not truly a scientist [pharmacist].

Knobe et al. take dual-character concepts to be a fundamentally new kind of concept, ones that they take to differ from other concepts in their underlying structure (see Fig. 1a). Because of this, they also wanted to compare dual-character concepts to "natural kind" concepts, which are assumed (at least in their original paper) to have a different underlying structure from dual-character concepts (see Fig. 1c).[16] Accordingly, they also included cases that described something that had some of the descriptive properties typically associated with that natural kind but lacking the relevant (alleged) underlying "essence" that people also associate with that natural kind. To illustrate, consider the case of 'zebra' described below:

> **'Zebra' description**: Jill loved going to the local zoo. Her favorite part of the zoo was the zebra enclosure. Since it was a small zoo, the enclosure had only one inhabitant. It had beautiful black and white stripes which Jill found mesmerizing. One day when the zebra got sick, the doctor began running tests on it and found that its DNA was unlike that of any previously studied zebras. Instead, the DNA was identical to that of a breed of donkeys.

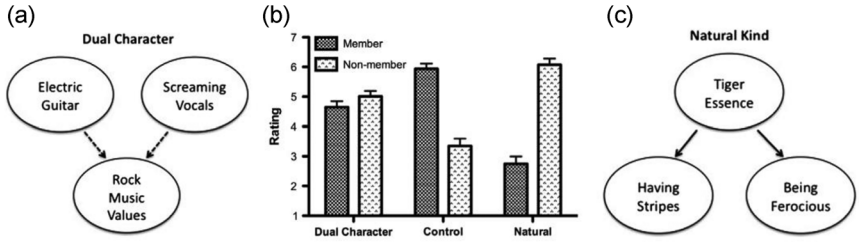

FIGURE 1 a) Proposed underlying structure of dual-character concepts, where the descriptive features (two top circles) are different ways of realizing an abstract value (bottom circle). b) Patterns of average truth-value judgments for "dual-character", "control", and "natural kind" terms from Knobe et al. Study 3. c) Proposed underlying structure of natural kind concepts, where the underlying essence (top circle) causes the observable descriptive features (bottom circles).

Here too, participants were asked to rate the truth of the relevant member and non-member assertions.

The pattern of judgments that Knobe et al. observed can be seen in Figure 1b. For terms (purportedly) tied to dual-character concepts, participants gave just above midpoint ratings for both the "member" and "non-member" assertions; for those terms (purportedly) tied to "control" concepts, participants gave high ratings to the member assertion and low ratings to the non-member assertion; and for those (purportedly) tied to natural kind concepts, participants gave low ratings to the member assertion and high ratings to the non-member assertion.

Knobe et al. take these patterns of judgments to provide evidence for the existence of a different kind of concept. The way they see it, natural kind concepts are meant to have an internal structure (depicted in Fig. 1c) whereby a given entity counts as falling in the extension of a particular natural kind concept *only* if the entity has the relevant underlying essence. Therefore, when it is lacking that essence (as in their descriptions), participants should disagree with the member assertion and agree with the non-member assertion. "Control" concepts are meant to have a different structure, whereby a given entity will fall in the extension of the relevant concept *only* if it meets the relevant descriptive criteria. Therefore, when those are satisfied (as in their descriptions), participants should agree with the member assertion and disagree with the non-member assertion. In contrast to these two "single-character" concepts, dual-character concepts are meant to have a novel kind of internal structure (depicted in Fig. 1a). This structure allows for a given entity to fall in the extension of a particular concept *either* if it satisfies the descriptive criteria *or* if it satisfies the abstract value that the descriptive criteria aim to realize. Therefore, when the descriptive criteria are satisfied but the normative criteria are not (as in the descriptions Knobe et al. used), participants should somewhat agree with both the member assertion and the non-member assertion. To make things easier going forward, we will refer to this pattern of judgments (similar overall agreement with both the member and non-member assertions) as "dual-character linguistic behavior", or just "dual-character behavior" for short.

As we understand it, the existence of dual-character concepts is meant to *explain* the relevant patterns in the linguistic data (including, crucially, the presence and absence of dual-character behavior), and to do so more successfully than alternative theories. For our purposes here, we can abstract away from some of the recent debates about what exactly the internal structure of dual-character concepts consists in, and instead focus on the general idea of what such concepts are meant to be. Put roughly, theorists in favor of the idea of dual-character concepts agree on the

following: some concepts (aka "dual-character concepts") have two kinds of criteria for determining what falls in their extension, one normative and one descriptive, in virtue of having a distinctive kind of internal structure. Importantly, in making this claim, theorists aren't just saying that there are certain patterns in linguistic behavior, but using new, technical language to do so. Rather, they are putting forward a theory about what explains the relevant kind of linguistic data. It's the fact that certain concepts have those distinctive characteristics, the thought goes, that explains the relevant patterns in linguistic data.

Before moving on, it's worth noting that it is not just Knobe, Prasada, and Newman who give this kind of "inference to best explanation" argument for the existence of dual-character concepts. Rather, this is the main *kind* of argument that all recent proponents of the existence of such concepts give, regardless of their further views on what exactly the "internal structure" of dual-character concepts is. Furthermore, even though theorists have subsequentially marshalled further linguistic data to support the existence of dual-character concepts, the primary *kinds* of linguistic data they use are tied to the linguistic phenomena that Knobe, Prasada, and Newman originally discussed.

3 | THREE NEW STUDIES, AND HOW THEY MATTER FOR THE MAIN ARGUMENT FOR DUAL-CHARACTER CONCEPTS

We now turn to the three new studies we conducted, and how they put pressure on the main argument for dual-character concepts we just discussed. The three studies we report suggest that the linguistic patterns to be explained are more complicated than proponents of dual-character concepts have claimed, and may in fact be better explained by appealing to general features of language use. Our first study finds that there are plenty of cases in which the terms that were originally taken to be associated with *non*-dual-character concepts can elicit dualcharacter linguistic behavior. Our second study further demonstrates that this same flexibility can be found for the terms that *were* originally linked to dual-character concepts: in some cases, they do elicit dualcharacter behavior, while in others they do not. Finally, a third study demonstrates how the broader conversational context can change whether a given term elicits the relevant linguistic behavior, even using the exact terms and descriptions that originally led Knobe, Prasada, and Newman to posit the existence of dual-character concepts.

Along the way, we also take a closer look at the original patterns that motivated Knobe, Prasada, and Newman to posit dual-character concepts. There we find that although we can indeed replicate the summary patterns they used to motivate positing dual-character concepts, their original way of summarizing the data obscured the fact that dual-character concepts typically do not *individually* elicit dual-character behavior. Instead, our data suggest the patterns exhibited in cases involving (purported) dual-character concepts are pretty straightforwardly continuous with the patterns exhibited in cases that involve (purported) non-dual-character concepts.

After putting these new data on the table, we put forward the outlines of a plausible alternative way of accounting for the relevant empirical patterns. While our aim here is not to offer a full positive account, we do think it's dialectically important to point out that there are other ways of explaining the relevant patterns. Our idea, in short, is to draw on already existing resources that philosophers and linguists use to explain variation and context-sensitivity in language use and show how these resources might well explain the relevant patterns originally associated with dual-character concepts. Once we appreciate these patterns in connection with a careful understanding of what dual-character concepts are meant to be, and what explanatory roles they are meant to

play in our thought and talk, it puts significant pressure on the motivation for positing that some concepts are in fact dual-character ones. Indeed, as we discuss later on, so too does it put pressure on the very idea of there being distinctive "dual-character" linguistic patterns associated with certain terms, regardless of whether one wants to posit certain concepts (at the level of thought) tied to those terms.

3.1 | Study 1: Flexibility in the elicitation of dual-character behavior using (purported) non-dual-character terms

In our first study, we wanted to see whether there are cases in which people use terms tied to (purported) non-dual-character concepts in a way that involves the patterns that Knobe, Prasada, and Newman associate with dual-character concepts. To get the basic idea, consider a term like 'second cousin' (which Knobe et al. take to be tied to a non-dual-character concept). Let's suppose (as Knobe et al. do) that our use of the term involves certain application-conditions, which, for our purposes, we can model as meeting certain criteria.[17] We can then test how people respond in cases where some of the (at least prima facie relevant) criteria are met, but others are not, and where we try to highlight criteria that might be particularly relevant in the context at hand. We can then ask: how similar do the patterns in these cases look to those that Knobe, Prasada, and Newman claim are best explained by positing that the term is tied to a dual-character concept?

To do all of this, we started by taking the original materials for the "control" concepts from Knobe et al., and then created additional contexts that highlighted the way in which the relevant criteria were only partially satisfied. While Knobe et al. only used materials involving 10 of the terms that they listed as control concepts that do not have a dual-character, we tested all 20.[18] To illustrate, consider 'table of contents' in the context below:

> **'Table of contents' description**: Laura is attempting to translate a collection of scientific essays that was originally written in Chinese into English. However, she is not particularly skilled at reading or writing in Chinese. In constructing the collection, not only did Laura mistranslate and misorder many of the section titles, she even misunderstood the book and inserted new essays that don't exist. The Table of Contents that she wrote does not at all accurately reflect what is included in the collection of essays. Further, it appears that she made several mistakes in the page numbers, so that the page numbers given in the Table of Contents do not direct the reader to the intended page in the book.

In the context of this case, consider both the member and the non-member assertions:

> **Member assertion**: There is a sense in which Laura has written a table of contents.

> **Non-member assertion**: Ultimately when you think about what it really means to write a table of contents, you would have to say that Laura did not truly write one.

What we wanted to investigate is whether participants' agreement with these two assertions would exhibit the key linguistic behavior that Knobe, Prasada, and Newman associated with purported dual-character concepts, even when all the terms used are linked to purported non-dual-character concepts.

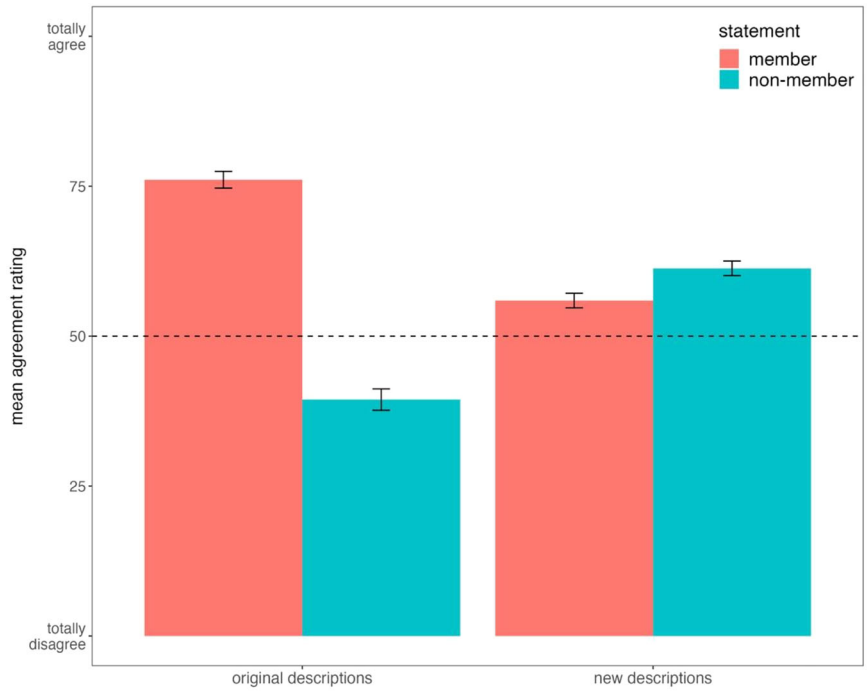

FIGURE 2 Participants' mean agreement rating with the member assertion (red bars) and non-member assertion (blue bars) for terms that are not meant to be linked to dual-character concepts. Left two bars depict agreement ratings in the original cases created by Knobe et al.; right two bars depict agreement ratings in the new cases we created. Error bars depict +/- 1 SEM.

To do this, we collected data from 100 participants ($M_{age} = 45.47$; $SD_{age} = 13.65$; 44 females, 56 males), who were presented with 15 different (purported non-dual-character) terms. After reading the relevant description (e.g., the 'table of contents' vignette above), participants rated their agreement with both the member and non-member assertions on a scale from 0 ('totally disagree') to 100 ('totally agree'). For some of the terms, participants read the original descriptions generated by Knobe et al., while for others, they read the new descriptions we created (as in the case of 'table of contents' above).[19]

Participants' agreement ratings with the two assertions indicate that we both replicated Knobe et al.'s original patterns using the original descriptions, and also that using our new descriptions, we were able to clearly elicit dual-character linguistic behavior (Figure 2).[20]

We can statistically quantify the observed patterns in two ways. First, we can ask whether the pattern observed for the member and non-member assertions differs between the original and the new descriptions (i.e., if there is an interaction effect between the kind of description and the agreement patterns for the two assertions). We found clear evidence for this interaction, $X^2(1) = 9.673$, $p < .001$.[21] Second, we can more specifically investigate whether this interaction effect occurred because participants' agreement exhibited a dual-character behavior in the new cases we created. Recall that the key characteristic of dual-character linguistic behavior is participants' willingness to agree overall with both the member and the non-member assertion (indicating two different kinds of criteria along which membership can be evaluated). We can statistically investigate whether this behavior was exhibited by testing whether the participants' agreement ratings

for both assertions were significantly above midpoint agreement in the newly created cases. We found that this was indeed the case for both the relevant member assertion ($M = 55.37$), $t(78) = 2.88$, $p = .005$, $d = .325$, and the relevant non-member assertion ($M = 61.63$), $t(78) = 6.10$, $p<.001$, $d = .687$. At a more qualitative level, this pattern is notably similar to that observed by Knobe et al., which in turn lead them to posit the existence of dual-character concepts (see Fig. 1b).

These results demonstrate that there seems to be a great deal of flexibility in *whether* a given term elicits dual-character behavior. Proponents of the existence of dual-character concepts might, however, be relatively unfazed by these data. After all, a number of them admit that, in certain contexts, one can get terms that are not associated with dual-character concepts to exhibit dual-character behavior. However, they usually insist that such cases involve contexts that are highly abnormal and so are not theoretically illuminating.[22] We don't think that these data can be dismissed so easily. This is partly because we don't see a clear reason to think that these contexts are any more abnormal than some of the key original ones involving, e.g., a person who conducts scientific studies and analyzes data but does not care at all about finding the correct answers. Moreover, we demonstrated this linguistic pattern across 20 terms that were assumed to be linked to non-dual-character concepts, while the original studies only demonstrated this pattern for 10 terms that were assumed to be linked to dual-character concepts. So, it seems that the cases we employed may be at least as easy to generate as ones involving purported dual-character concepts.

In this study, we used the terms that were originally linked to (purported) non-dual-character concepts, and we found that they both can elicit dual-character behavior (in our new cases) and can *not* elicit dual-character behavior (in the original cases). This suggests that dual-character behavior may not be best explained by a feature of the structure of the concept that the term is tied to, but instead may be better explained by some other aspect of language use. A further question is whether the same conclusion can also be extended to the dual-character behavior that was originally observed for terms that are purportedly linked to dual-character concepts.

If the dual-character behavior is indeed best explained by some general feature of language use, we might expect a similar level of flexibility for both terms that are and are not linked to purported dual-character concepts. We investigate this flexibility for terms that are linked to purported dual-character concepts next.

3.2 | Study 2: Flexibility in the elicitation of dual-character behavior using (purported) dual-character terms

In this next study, we sought to investigate whether terms associated with dual-character concepts exhibit similar flexibility in the extent to which they elicit the key pattern of linguistic behavior on member and non-member assertions. To pursue this question, we took Knobe, Prasada, and Newman's' original materials for their dual-character cases and then created a second set of descriptions in which there are again normative criteria that are not satisfied, but where these normative criteria are not clearly relevant to the way the term is being used. Knobe et al. originally only tested 10 terms, while we tested all 20 of the terms linked to their original list of 20 dual-character concepts.[23] To illustrate our new cases, consider the description of the person below, and the term 'soldier':

> **'Soldier' description**: Darryl is currently on deployment for the United States Military. Darryl receives payment and regular benefits for his work in the military and is expected to be on deployment for another 3 months with his team. However,

according to his fellow soldiers and commanding officers, he is a difficult soldier to work with. He has poor directional skills, takes criticism badly, and often picks petty fights with his fellow soldiers, which his officers say sows distrust within the squads.

With this case in mind, now consider the two following assertions:

Member assertion: There is a sense in which Darryl is a soldier.

Non-member assertion: Ultimately when you think about what it really means to be a soldier, you would have to say that Darryl is not a soldier.

What we wanted to investigate is whether participants' agreement with these two assertions would cease to exhibit the key linguistic behavior that Knobe et al. associated with purported dual-character concepts when the normative criteria that are not met are not clearly relevant for the use of the term.

We collected data from 100 participants ($M_{age} = 41.90$; $SD_{age} = 14.00$; 41 females, 57 males, 2 non-binary), who were presented with 15 different (purported dual-character) terms. After reading the description, participants rated their agreement with both the member and non-member assertions on a scale from 0 ('totally disagree') to 100 ('totally agree'). In addition, we asked participants to rate the extent to which they perceived the entity described as "atypical" and "non-ideal" to ensure that we succeeded in creating somewhat similar contexts in which the entity described did not satisfy normative criteria that are associated with the relevant term. For some of the terms, participants read the original descriptions generated by Knobe et al., while for others, they read the new descriptions we created (as in the case of 'soldier' above).[24]

Participants' agreement ratings with the two assertions indicate both that we replicated Knobe et al.'s original patterns using their original descriptions and that, in our new cases, we were able to clearly elicit the linguistic behavior previously associated with non-dual-character (or "control") concepts (see Figure 3).

As a first step in a more rigorous analysis of the data, we wanted to statistically verify that our new descriptions succeeded in creating cases where participants would agree that the entity described was "non-ideal" and "atypical". Our "manipulation check" questions allowed us to verify that this was the case, as participants judged the entity described in the new descriptions to be both "atypical" ($M = 59.92$), $t(80) = 5.827$, $p<.001$, $d = 0.647$ and 'non-ideal' ($M = 60.88$), $t(80) = 5.90$, $p<.001$, $d = 0.655$.[25] This confirms that the new descriptions we created were somewhat similar to the original cases in that they also involved normative criteria that were not satisfied.

Next, we wanted to statistically quantify the observed pattern in two key ways. The first is to ask whether the pattern observed for the member and non-member assertion differs between the original and the new descriptions (i.e., if there is an interaction effect between the kind of description and the agreement patterns for the two assertions). We found clear evidence for this interaction, $X^2(1) = 17.730$, $p < .001$.[26] Second, we wanted to more specifically investigate whether this interaction effect occurred because the terms purportedly associated with dual-character concepts elicited a non-dual-character pattern of responses with the new descriptions. Given that the new descriptions were ones in which the normative criteria were meant to *not* be relevant to the use of the term, the key non-dual-character pattern would be participants' willingness to agree with the member assertion but disagree with the non-member assertion. In line with these predictions, we found that participants' agreement ratings with the member assertion were significantly

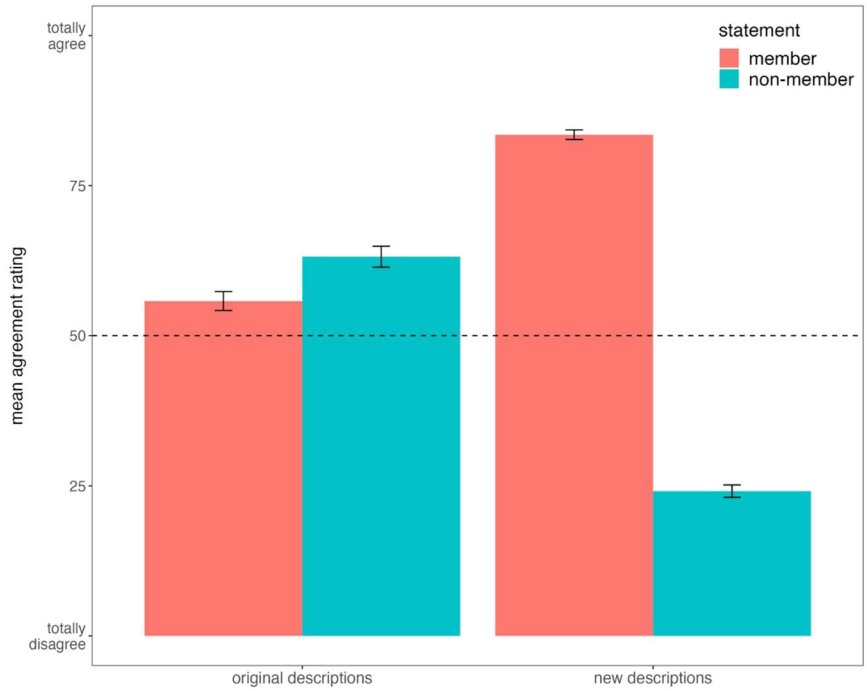

FIGURE 3 Participants' mean agreement rating with the member assertion (red bars) and non-member assertion (blue bars) for terms that are purportedly linked to dual-character concepts. Left two bars depict agreement ratings using the original case descriptions by Knobe et al.; right two bars depict agreement ratings using the new descriptions we created. Error bars depict +/- 1 SEM.

above the midpoint of the scale in the new cases, $t(80) = 30.24$, $p<.001$, $d = 3.359$. At the same time, their agreement with the non-member assertion was significantly *below* the midpoint in the new cases, $t(80) = -18.47$, $p<.001$, $d = 2.052$. At a more qualitative level, the pattern observed for terms which Knobe et al. took to be linked to dual-character concepts is notably similar to the pattern they originally found for the purported non-dual-character cases (see Fig. 1b).

In short, we sought to investigate the flexibility of dual-character behavior for terms purportedly linked to dual-character concepts. We found that such terms are highly flexible in terms of whether or not they elicit dual-character behavior in a way that seems plausibly similar to terms that were associated with (purported) non-dual-character concepts.

To return to our earlier dialectic, we initially observed that terms purportedly linked to non-dual-character concepts exhibit dual-character behavior in some cases, but do not in others. This initial finding suggested that dual-character behavior may not be best explained by a feature of the structure of the concept that the term is tied to, but instead may be better explained by some other aspect of language use. We then sought to ask whether the same pattern of flexible dual-character behavior could be found for the terms that were purportedly linked to dual-character concepts. If dual-character behavior is indeed best explained by some general feature of language use, we should expect a similar sort of flexibility for terms that are linked to (purported) dual-character concepts. As described above, we found clear evidence for such flexibility.

One response that proponents of dual-character concepts may be tempted to give at this point is to argue that what is special about dual-character concepts (and the associated linguistic behavior)

is that they will not exhibit such flexibility when the *right* normative criteria have been targeted. That is, perhaps our new studies were only able to demonstrate flexibility for dual-character concepts because our new descriptions did not target the relevant abstract values for the dual-character concept concepts, and thus it is unsurprising that in such cases people judged the term to apply to the entity described.[27]

Knobe et al. are unclear about how exactly we should figure out what the relevant normative dimensions of dual-character concepts are (i.e., the abstract values that the descriptive features generally help to realize). They suggest that the relevant abstract values are tied to judgments about something being a "true X" (vs. a "good X"), but it's not even totally clear how they were able to make sure that *they* targeted the relevant abstract values in their original cases. There are often many different abstract values that could be realized by the descriptive features associated with a concept or term. For example, the concrete descriptive features often associated with SOLDIER (fighting in a war, being part of an army, obeying commands, using weapons, etc.) are plausibly ways of realizing many different abstract values such as fraternity, loyalty, bravery, self-sacrifice, or even self-reliance. In our new description for 'soldier' (included above), we used an example where the person described was clearly lacking in the abstract values of fraternity, loyalty, and also perhaps self-reliance. Of course, it is possible that these are simply not the *right* abstract values that are the relevant part of the dual-character concept SOLDIER. That is: perhaps they are abstract values tied to some way(s) we assess soldiers, but not values that are built into the relevant part of the (purported) dual-character structure of SOLDIER. We think this kind of response highlights that extant theories of dual-character concepts are worryingly underspecified about key details regarding which specific abstract values are built into the relevant part of specific dual-character concepts, even when they have given overall theories about which abstract values matter, and why. Yet we grant that it may very well be the case that our new descriptions did not manipulate the key abstract values that proponents of dual-character concepts had in mind for these particular concepts.

One way to address this response is to simply stick to the descriptions that Knobe et al. originally created, since we can be relatively certain that those cases at least target whatever abstract values Knobe et al. had in mind for each purported dual-character concept. If the best explanation for dual-character behavior is some general property of language use, we might still expect to find flexibility in the extent to which these exact descriptions elicit dual-character behavior depending on the broader conversational contexts in which those descriptions occur.[28] We take up this further question next.

3.3 | Study 3: Context-based flexibility in the elicitation of dual-character behavior

We took the original materials for the 10 dual-character concepts that Knobe et al. tested and created two larger conversational contexts in which each of the original descriptions involving those concepts occurred.[29] To illustrate, consider the original description used by Knobe et al. when testing the term 'artist':

> **Original description**: Greg makes a living by creating paintings, which then appear in art galleries and museums. However, Greg does not actually care at all about the aesthetic value or originality of the paintings he creates. His only goal is to make lots of money, and he does that by carefully keeping track of the latest fashions in the art world and putting together paintings that fit these fashions exactly.

We then asked participants to respond to this description as embedded in two different contexts. The first was designed such that salient normative criteria in the context were plausibly relevant to the way the term is being used:

> **Relevant criteria context**: At a café, a group of different kinds of artists are gossiping about other artists and someone asks whether Greg counts as an artist.

The second was designed such that the same normative criteria were (at least prima facie) less salient in the context and less relevant to the way the term is being used:

> **Irrelevant criteria context**: A firm of tax attorneys have just gotten a bunch of new clients, including Greg, and they need to figure out how to divide them up. Some of the attorneys specialize in small family farms, some specialize in corporate taxes, and some specialize in the tax laws that apply to artists and musicians. They are discussing which tax attorney should handle Greg's case and so need to decide if he counts as an artist.

With each of these cases in mind, we now want to consider both the member and non-member assertions. Here, we use the simple member and non-member assertions used in Experiment 4 of Knobe, Prasada, and Newman's original 2013 paper, as follows:

> **Member assertion**: Greg is an artist.

> **Non-member assertion**: Greg is not an artist.

We used these simplified member and non-member assertions because, as pointed out by Knobe et al. in their original paper, the form of the non-member assertion previously used explicitly highlights the relevance of certain normative dimensions by including language such as 'ultimately' and 'really', and we didn't want these terms to reintroduce the relevance of these normative dimensions in the irrelevant criteria context.[30]

To test whether we continued to observe flexibility in dual-character behavior based on the broader conversational context, we collected data from 100 participants (M_{age} = 39.27; SD_{age} = 13.06; 48 females, 51 males, 1 non-binary), who were presented with 10 different terms (purportedly linked to dual-character concepts). After reading the original description, followed by either the relevantcriteria conversational context or the irrelevant criteria context, participants rated their agreement with the member and non-member assertions on a scale from 0 ('totally disagree') to 100 ('totally agree').[31]

Participants' agreement ratings indicate that we replicated Knobe et al.'s original patterns they observed in Study 4 in the relevant criteria contexts.[32] However, we also see evidence of clear non-dual-character behavior in the irrelevant criteria contexts (see Figure 4).

Once again, we can statistically quantify the observed patterns in two ways. The first is to ask whether the pattern observed for agreement with the member and non-member assertions differs between the relevant criteria and irrelevant criteria contexts (i.e., if there is an interaction effect between the kind of context and the agreement patterns for the two assertions). We found clear evidence for this interaction, $X^2(1) = 4.293$, $p < .001$.[33] Second, we wanted to more specifically investigate whether this interaction effect occurred because we observed a dual-character pattern of responses in the relevant criteria contexts, but a non-dual-character pattern of responses in

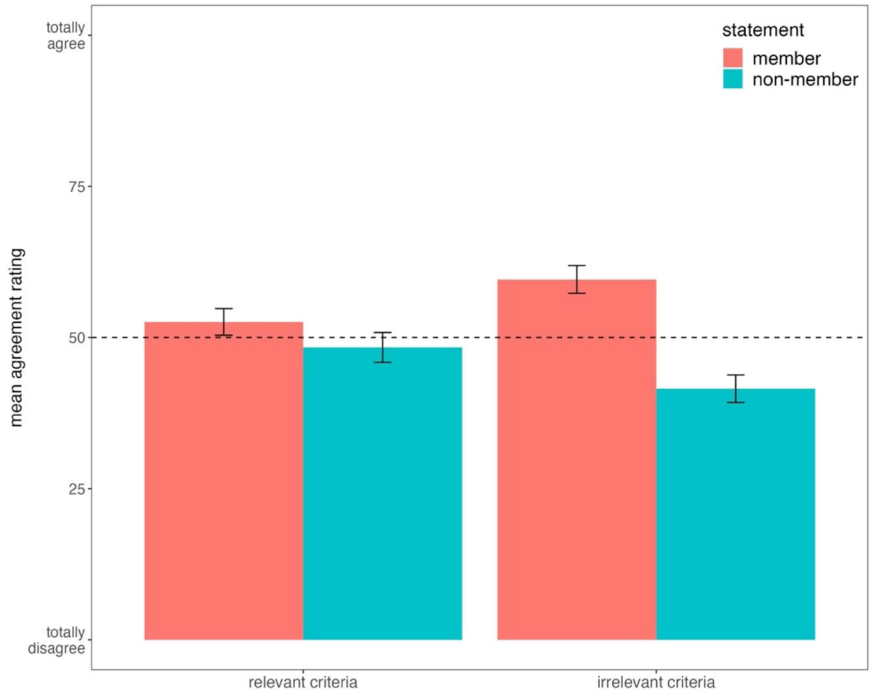

FIGURE 4 Participants' mean agreement rating with the simplified member assertion (red bars) and non-member assertion (blue bars) for the 10 terms that are purportedly linked to dual-character concepts. Left two bars depict agreement ratings in contexts where the identified normative criteria are relevant; right two bars depict agreement ratings in contexts where these normative criteria are irrelevant. Error bars depict +/- 1 SEM.

the irrelevant criteria contexts. The patterns were broadly in line with these predictions. In the relevant criteria contexts, we found that participants' overall agreement with both the member assertion ($M = 54.39$) and non-member assertion ($M = 49.41$), did not differ significantly from the midpoint of 50 (p's > 0.14), and did not differ significantly from each other, $t(151.14) = 1.256$, $p = 0.211$, $d = 0.202$. This midpoint agreement with both assertions is slightly higher than, but otherwise similar to, the pattern observed in Experiment 4 in Knobe et al.'s original 2013 paper (where they also employed the same simpler member and non-member statements we did here).[34] By contrast, in the irrelevant criteria contexts, we found that participants' overall agreement with the member assertion ($M = 58.82$) was significantly *above* the midpoint ($t(77) = 2.869$, $p = 0.005$, $d = 0.325$), while agreement with the non-member assertion ($M = 41.50$), was significantly *below* the midpoint ($t(77) = -3.057$, $p = 0.003$, $d = 0.346$). Unsurprisingly, then, participants' agreement ratings with the two assertions differed significantly from each other, $t(152.49) = 4.179$, $p < 0.001$, $d = 0.669$.[35] That is, when the normative criteria that were not satisfied were not relevant to the use of the term in the conversational context, we observed a pattern that Knobe et al. associated with non-dual-character (or "control") concepts.

In sum, the pattern we find is one where we observed dual-character behavior when the originally-targeted normative criteria were relevant in the conversational context, but also lacking that behavior when these normative criteria were irrelevant in the broader conversational context. Critically, we observed this contextual variation in whether dual-character behavior was exhibited using the same terms and same descriptions as in the original studies, so we

can be relatively certain that we targeted the intended normative dimensions (whatever those might be).

3.4 | Summing up

Let's take stock. The main motivation for positing the existence of dual-character concepts has been to explain a particular pattern of linguistic behavior. Knobe, Prasada, and Newman (and others following them) use an inference to best explanation argument for why we should believe in dual-character concepts, and for why we should believe that certain concepts are dual-character ones while others are not. The three studies we have reported here suggest that the linguistic patterns to be explained are much more complicated than originally suggested. Collectively, the three studies provide increasing evidence that the relevant linguistic behavior is likely better explained by general features of language use.

Specifically, the results of our first study found that terms that were originally associated with non-dual-character concepts also elicit dual-character behavior in some cases, while they do not in others. The results of our second study found a similar kind of flexibility for the terms that were originally meant to be linked to dual-character concepts: in some cases, they elicit dual-character behavior, while in others they do not. And the results of our third study demonstrated that the broader conversational context can change whether a given term elicits dual-character behavior, even when we use the exact terms and descriptions that originally led Knobe, Prasada, and Newman to posit the existence of dual-character concepts.

4 | A POSSIBLE DIFFERENT KIND OF EXPLANATION

Throughout our discussion, we've gestured toward the idea that the relevant linguistic data might be best explained by relatively general facts about how people use language, rather than by the presence of a special type of concept involved in their thinking. We now want to turn to what such an explanation might look like. Our goal in discussing this alternative kind of explanation isn't to try to establish that it is obviously correct. Rather, it is to advertise what we take to be some of its main merits, and, in so doing, put more pressure on the idea of positing the existence of dual-character concepts.

Here's a possible different kind of explanation of the relevant linguistic data we've been discussing. Take a given word (e.g., 'scientist') or conventionalized combination of words (e.g., 'rock music') that Knobe et al. associate with a concept (e.g., SCIENTIST or ROCK MUSIC). We can note that there are a range of ideas and inferential patterns often closely associated with the use of that word (or conventionalized combination of words). Some of those things might help compose the "meaning" of that word (or with that conventionalized combination of words) – in at least some good sense of "meaning" that we should care about. For example, some might be part of the kind of "meaning" that constitutes the lexically encoded information which linguists seek to capture when doing semantics. (And, plausibly, given the diverse intellectual traditions within semantics, there might be multiple different things different linguists are trying to capture here). Others might just be closely associated with the use of that word (or with that conventionalized combination of words), but not, strictly speaking, part of the meaning (e.g., they may involve pragmatic inferences based on the way a word is used in a given context). With this basic idea about context-sensitivity in hand, we suggest the following: in a given conversational context, certain of

those ideas and inferential patterns will be more or less salient to the speakers in that context, and certain of those things will be ones that speakers care more or less about in that context. Both of these things plausibly have to do (at least partly) with speakers' non-linguistic understandings of the entity in question. But they plausibly also have to with other factors too, such as the stakes of what is involved in the conversation and the practical aims of the speakers. If that is right, then it might be that part of what is going on in our cases is that the descriptions and conversational contexts are shifting which of the ideas and inferential patterns participants put the most weight on for determining whether to apply a term or not.

To briefly flesh out how this alternative proposal is meant to work, consider the example of judgments involving the term 'artist' in our third study. In the initial description, participants were told (among other things) that a man named Greg creates artwork that is hung in museums and galleries and that this is one way he aims to make money. They are also told that Greg does not care about the aesthetic value or originality of the art he creates. On this alternative proposal, each aspect of the description makes salient at least one feature that may be relevant to the use of the term 'artist'. After all, each feature was one that the speaker (whoever that is) deemed worth mentioning in describing Greg. In our third study, recall that this description was then embedded in one of two larger conversational contexts. In one context (the relevant criteria context), a group of artists is critically evaluating other artists. Given the conversational context, it is natural to assume that whether a person cares about the aesthetic quality of their work and whether they seek to create something original are salient features (and ones that these artists will also care about) that will guide them in deciding whether the term 'artist' applies to Greg. As Greg is described as lacking these features, participants would likely be willing to endorse the assertion "Greg is not an artist" because Greg fails to satisfy at least some of the criteria most salient (and most valued) in this conversational context. By contrast, in the second context (the irrelevant criteria context), a group of tax attorneys is trying to divide up clients based on the different attorneys' expertise. Given this conversational context, it is natural to assume that whether the person cares about the aesthetic quality of their work and whether they seek to create something original are not salient features that will guide the lawyers use of the term 'artist', nor something any of these lawyers particularly care about for applying the term in this context. For better or for worse, such features do not have any bearing on which tax laws apply to which individuals. Instead, the relevant criteria that guide the use of the term 'artist' in this context are presumably features such as whether the person makes an income selling paintings, and Greg clearly does satisfy this criterion. Accordingly, in the second context, if participants are sensitive to the aims of the participants in that conversation, they should be willing to endorse the assertion that Greg is an artist, and not the assertion that Greg is not an artist, which is what we observed.

In proposing this alternative explanation of the linguistic data, we've only given the outline of a general idea. There will be different ways of making it more precise. For example, one way to do so would be to embed it in a theory of what composes the meaning of the relevant words, which distinguishes what's in the semantics of the relevant words from the pragmatics of how we use them (e.g., involving such things as implicature). Such a theory might involve not only taking a stand on which kinds of things in general compose the meaning of words (e.g., application-conditions, inferential patterns, etc.) but also what the specific versions of those things are for the words involved in the relevant studies (e.g., what the application-conditions are that constitute the meaning of the word 'scientist'). Another way to make our general proposed explanation more precise would be to take a stand on which different shifts in context yield different results in which inferential patterns or ideas closely associated with a given term speakers give more importance to, and why. These two ways of making the proposal more precise would interact in key ways,

including whether to describe certain shifts in how speakers use words as shifts in "meaning". For our purposes here, what's important is not how such details are best worked out (which we hope those sympathetic to this line of thought will attempt to do in the future). Rather, what's important is just that this is an alternative line of explanation for the relevant linguistic data, which, at the very least, deserves to be taken seriously and further explored.

One thing we want to underscore about this kind of explanation is that it fits naturally with general views about language use that emphasize linguistic flexibility across contexts. This is especially so if those views emphasize ways that people use language in ways that are sensitive to the varied aims, audiences, and information people have across different contexts. For example, take the idea of "metalinguistic negotiation" (or, equivalently, "normative metalingiuistic disputes") put forward by one of us (Plunkett) in co-authored work with Tim Sundell.[36] Put roughly, in metalinguistic negotiations, speakers implicitly put forward, via different "metalinguistic" usages of a term, conflicting normative views about how a given term (e.g., 'scientist') should be used. The speakers might endorse conflicting views about what the fundamental meaning of the word in question should be, or they might have a view on something less foundational, such as how they should set the threshold for a context-sensitive term with a threshold parameter. We aren't here endorsing this model of metalinguistic negotiation, let alone the idea that it happens regularly in discourse. Rather, we are just noting that the account we are sketching here has an affinity with the idea of metalinguistic negotiation, which involves the idea that speakers can use language in a way where speakers can contest different inferences or ideas associated with a term, and attempt (plausibly at least *sometimes* successfully) to shift them in different ways. If that kind of theory has independent support, it might well bolster the sort of view we are floating here about how to best explain the linguistic data discussed in this paper.[37]

Furthermore, it's worth noting that one of us (Plunkett), in co-authored work with Sundell and Rachel Sterken, has recently argued that the phenomenon of metalinguistic negotiation might be a part of the best explanation of some features of how we use certain generic statements (e.g., "boys don't cry", as uttered in certain contexts) in some cases where others (e.g., Leslie) have appealed to Knobe to dual-character concepts.[38] This argument suggests a general strategy which would extend our work in this paper in a more ambitious way. In short, the idea is to appeal to a range of facts about how speakers modulate and shift how they use words across contexts (including facts about how speakers engage in metalinguistic negotiation), as well as to other general features of thought and talk that we already have reason to believe in, to explain *all* the key linguistic patterns that theorists have thought we need dual-character concepts to explain. Whether or not that can be done, of course, takes us far beyond the scope of this paper. Here, we want to suggest something much more modest: namely, that the relevant linguistic data that Knobe et al. draw our attention to, as well as the data we've added in our additional studies, might best be explained in the way we've outlined above, rather than by positing dual-character concepts or, indeed, new concepts of any kind.

In closing our discussion of this alternative kind of proposal, it is worth noting that the basic idea behind it – namely, that context can matter in significant ways for which inferential roles or ideas associated with a term speakers emphasize, and in ways that connect to the explanation of dual-character behavior – is one that Tobia, Newman, and Knobe also consider in recent work.[39] They, however, posit a more limited role for this idea, alongside emphasizing the (purportedly) distinct "dual-character" structure of how we use certain terms. What we are suggesting here is that this basic idea might do much more of the explanatory work than they suppose in accounting for the relevant linguistic data, and that, as a result, we may be able to dispense with positing dual-character concepts altogether.

4.1 | Examining variation in dual-character behavior

With this sketch of one alternative way of explaining the linguistic patterns in hand, we want to briefly address an important difference in the empirical predictions made by our account vs. Knobe, Prasada, and Newman's. On the proposal that dual-character linguistic behavior is caused by differences in the structure of the underlying concepts, we should expect to see substantial homogeneity in the extent and stability of dual-character behavior, at least for (purportedly) unambiguous cases of dual-character concepts. To get a sense for why, it may be helpful to return to the comparison to natural kind concepts, which Knobe et al. use to motivate the idea that some kinds of concepts have a special internal structure. Recall that natural kind concepts, such as RACOON or WATER, were proposed to have a "single-character", whereby a given entity counts as falling in the extension of that concept *only* if the entity has the relevant (alleged) underlying essence (see Fig. 1c). Accordingly, one can vary the features of something that falls in the extension of a natural kind concept that are tied to the descriptive criteria typically associated with it, but without altering whether that thing still falls in the extension of that concept. To illustrate: the idea is that people generally think that a racoon that is painted to look like a skunk, given a stinky pouch, etc., is still not a skunk.[40] For this kind of special class of concepts that (purportedly) share a common kind of internal structure, we should expect that the associated terms will elicit relatively homogenous linguistic behavior in cases where the entity in question lacks the (alleged) underlying causal "essence" (see Fig. 1). In other words, people generally think that a raccoon painted like a skunk is still a racoon (and not a skunk) for the same reasons and to roughly the same extent that a rock sculpted to resemble a person is still a rock (and not a person). In such cases, people should generally disagree with the member assertion and generally agree with the non-member assertions. In a similar way, proponents of dual-character concepts posit a special kind of internal structure that is shared across dual-character concepts. Thus, a natural prediction for such accounts to make is that there will be (at least somewhat) similar homogeneity when the entity in question meets the typical descriptive criteria associated with the dual-character concept but does not satisfy the abstract value that these descriptive criteria (by hypothesis) help realize. In such cases, people should somewhat agree with *both* the member and non-member assertions.

This prediction stands in stark contrast to the predictions of the alternative account sketched above, which appeals to quite general features of language use. This alternative account is based on the idea that there is a great deal of flexibility in determining which features are relevant for the way a term is being used in a given conversation, and on the suggestion that it is not always clear-cut which features are relevant; rather, these issues of relevance are often inferred or negotiated (whether implicitly or explicitly) in a conversation. In line with this flexibility, such accounts naturally predict substantial variation in participants' agreement with the member and non-member assertions depending on the particular features highlighted in the descriptions, perceptions of how the term is typically used, differences in inferences about what is relevant in a given conversational context, and differences in what speakers care about.

Given the differences in the expected variation predicted by these two accounts, we want to return to the data we have already collected to see if they can help adjudicate between these accounts. One straightforward place to investigate variation is in our data about Knobe et al.'s original cases: that is, the subset of our cases where we used their proposed dual-character terms and their original descriptions of entities that meet the associated descriptive criteria but do not satisfy the associated abstract value. Focusing on just these cases, do we find a homogenous pattern of responses where participants tend to somewhat agree with both the member and

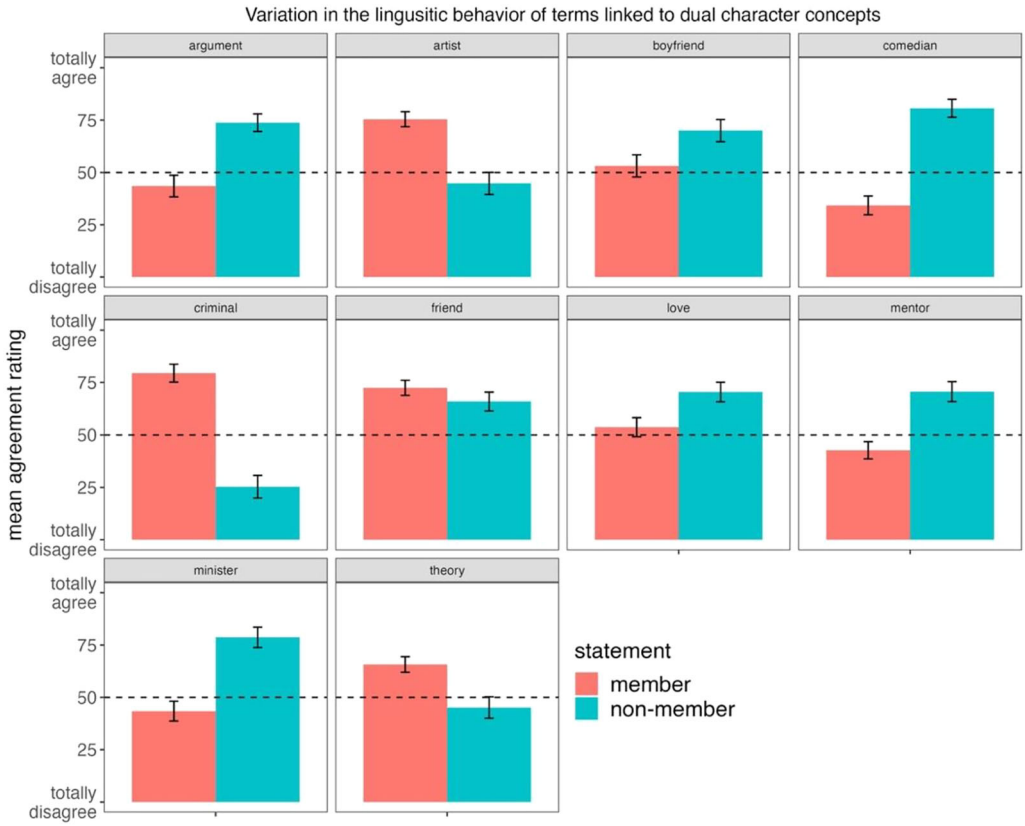

FIGURE 5 Participants' mean agreement rating with the member assertion (red bars) and non-member assertion (blue bars) separately for each term that was proposed to be linked to a dual-character concept, using only the original terms and descriptions from Knobe, Prasada, and Newman. Error bars depict +/- 1 SEM.

non-member assertions? In short, no. What we find is a great deal of variation across the different terms (Figure 5).

For some terms, such as 'comedian', 'mentor', 'argument', and 'minister', it seems as though participants basically agreed that the term does not apply to the entity described (disagreeing with the member assertion and agreeing with the non-member assertion); for other terms, such as 'criminal', 'artist' and 'theory', it seems that participants straightforwardly *agreed* that the term applied to the entity described (agreeing with the member assertion and disagreeing with the non-member assertion); and for the remaining three terms, agreement with these two assertions fell at various points between these two endpoints. In fact, the linguistic pattern that we have been characterizing as "dual-character linguistic behavior" (agreeing with both the member and the non-member assertions) was only straightforwardly observed in the case of 'friend'. Overall then, it seems as though the very pattern of judgments originally posited as being distinctive of dual-character concepts (see Fig. 1b) is largely an artifact of averaging across various kinds of *non*-dual-character patterns.

In addition, we might want to take a broader look at the observed variation across all of the terms we used throughout our studies, by comparing the amount and kind of variation one finds for (i) the terms that were originally proposed to be associated with dual-character concepts and (ii) the terms that were instead proposed to be associated with non-dual-character concepts.

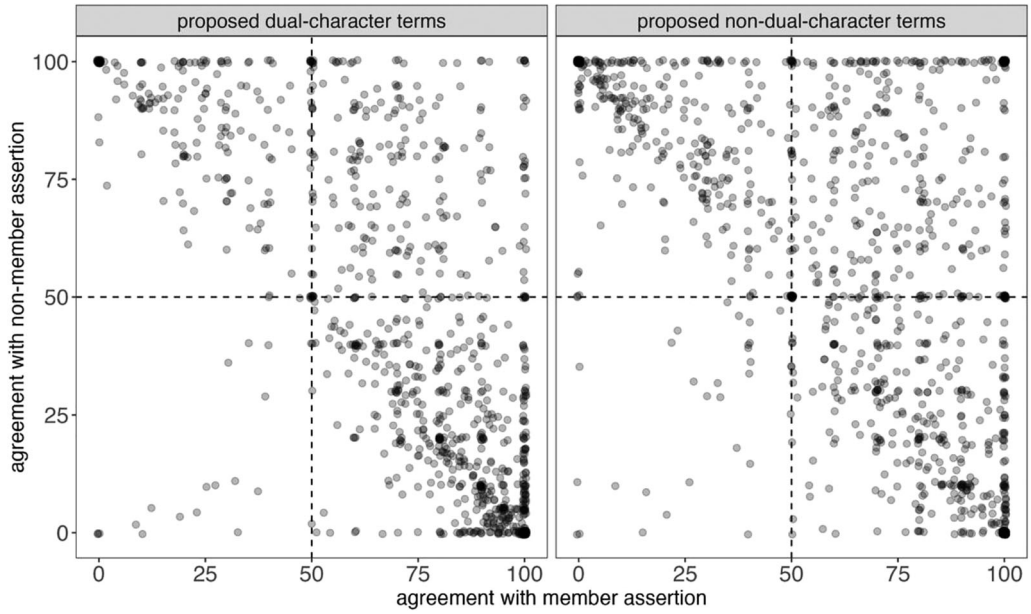

FIGURE 6 Individual participants' judgments from studies 1 and 2 with both the member assertion (x-axis) and non-member assertion (y-axis). The left plot depicts agreement ratings concerning only terms associated with purported dual-character concepts; the right plot depicts agreement ratings concerning only terms associated with purported non-dual-character concepts. Points located in the top right quadrant of each plot are patterns of judgments that reflect the dual-character linguistic behavior.

Accounts that appeal to general features of language use can allow for substantial variation in participants' agreement ratings for both kinds of terms, and so it is worth asking whether such variation is actually observed in our data. Comparing the data from the first two studies, indeed what we find is a great deal of variation in participants' responses concerning both the terms that were proposed to involve dual-character concepts and the terms that were not; we also find remarkably little difference in the patterns of variation between the two groups (see Figure 6).[41] In fact, to the extent that we do see differences in the distributions, it is because responses to terms linked to (purported) dual-character-concepts are more likely to show *non*-dual-character behavior. At a more quantitative level, what we find is that for proposed dual-character terms only 15.14% of the time did individual participants give responses that can be characterized as dual-character (these are found in the upper right quadrant of the left graph in Figure 6); in contrast, for proposed non-dual-character terms, 22.39% of the time participants gave responses that indicated a dual-character pattern (these found in the upper right quadrant of the right graph in Figure 6). Moreover, aggregating across all three of our studies, and considering the average pattern observed for each different case where a term was used, we find that for proposed dual-character concepts, there were only 6 cases (out of 50) where the average response indicated a dual-character linguistic pattern; in contrast, for proposed non-dual-character concepts, we found this dual-character pattern for 12 cases (out of 31).

In short, a closer look at the variation in agreement ratings with the member and non-member assertions does not fit well with accounts that seek to explain the observed linguistic patterns by appealing to differences in underlying conceptual structure. These accounts predict shared patterns of responses for dual-character concepts because they appeal to the shared structure of these

concepts to explain the overall pattern of data that was observed. However, a closer investigation of the data reveals that this overall pattern hides a great deal of variation even within the original group of proposed dual-character concepts (Fig. 5). By contrast, accounts that seek to explain the overall patterns in terms of general features of flexible language use do not predict homogeneity in the patterns of responses for different terms. Rather, they predict that a somewhat similar amount of variation could be found for both terms that were proposed to be linked to dual-character concepts and those that were not, and indeed, similar patterns of variation were in fact observed (Fig 6).

5 | THREE POSSIBLE RESPONSES ON BEHALF OF PROPONENTS OF DUAL-CHARACTER CONCEPTS, IN LIGHT OF OUR ARGUMENT

We now want to consider three possible responses to our main line of argument. While we don't think any of these responses are convincing, we think they all raise important theoretical issues that help clarify the relevant philosophical terrain and help underscore interesting lines for future research.

5.1 | A bifurcated explanation?

One response aims to bolster the case for dual-character concepts by proposing a kind of bifurcated explanation of the relevant linguistic data we've been discussing. Here's the basic idea: perhaps the dual-character behavior of how we use *certain* words is explained by the "dual-character" nature of the relevant concepts, whereas the dual-character behavior in our use of other words is explained just by general features of thought and communication (e.g., of the kind that we floated in our possible alternative explanation). One reason to take this idea seriously is that Knobe, Prasada, and Newman themselves suggest that general features of how we use language (in particular, our use of "hedges"[42] and how our use of language is related to "task variation"[43]) explain what's going on with sentences involving (by hypothesis) "control concepts" that might (at least prima facie) seem to exhibit certain patterns that are "dual-character"-esque.[44] Given this idea, perhaps one could make this claim more ambitious, to attempt to account for the full range of "dual-character"-type patterns we've discussed in this paper, involving what Knobe et al. take to be "control concepts".

It's possible that the best explanation of the relevant linguistic data is a bifurcated one along these lines, on which the linguistic behavior using certain words (those purportedly tied to dual-character concepts) gets one explanation, while the same behavior involving other words (those purportedly not tied to dual-character concepts) gets a totally different one. But we see little reason as of yet to accept that kind of explanation. To illustrate, to make a proposal convincing by appeal to certain general features of language use (e.g., the use of "hedging" or the importance of "task variation"), proponents of dual-character concepts would need to tell us more about what these features of language involve and what tests there are for showing that certain sentences involve such things. They would then need to apply these tests across the board, both to the cases involving purported dual-character concepts and to the cases involving purported non-dual-character concepts. Finally, it is worth pointing out that for this kind of bifurcated account to be successful, one would still need to separately explain why words tied to purported dual-character concepts sometimes seem to function in a way that patterns like words tied to purported non-dual-character

concepts (as our third study shows). We think this idea of a "bifurcated response" suggests an interesting line of thinking for future research. But, given the lay of the land right now, it certainly can't be taken for granted that things will vindicate a cleanly bifurcated explanation. One might also say something even stronger. Insofar as simplicity and unity are virtues in explanation, then (at least on a number of plausible ways of understanding such virtues), those things would seem to disfavor this kind of bifurcated account of what explains the relevant linguistic behavior.

5.2 | A lower bar for positing new "kinds" of concepts?

A second response to our skepticism about dual-character concepts begins from the idea that we are being too demanding in terms of our understanding of what it would take to be a special "kind" or "type" of concept. One way to develop this idea is as follows. One might think that we don't need to be able to locate a bright line between concepts whose deployment always leads to certain kinds of linguistic behavior and those whose deployment does not. Or at least one might think this is so in the context of the relevant kind of theorizing about thought and talk that the idea of "dual-character" concepts is meant to contribute to. Instead, one might think it would be enough to identify a certain theoretically salient threshold, and then to claim that certain concepts fall on one side of that threshold whereas others do not.

We think this too is an interesting response to explore. And it might well be that, at the end of the day, some appropriately subtle version of this idea is the best way to support the basic idea of dual-character concepts. But we also have reason to doubt its prospects for success. To see why, consider the following question: what, exactly, would the relevant threshold be? Given the variation and context-dependence we've found with the relevant linguistic behavior, it's hard to see how to set the bar in a theoretically principled way, which is responsive to the kind of explanatory work that the existence of dual-character concepts is meant to do.

To push this idea, consider that, as we have seen, it seems likely that there are interesting patterns in how often, when, and why people use terms in a way that exhibits the relevant kind of dual-character behavior. We don't doubt that there are theoretically useful cuts to be made that help explain aspects of those patterns. But once we start digging into those patterns, we can see that there are a number of differences we can wonder about. For example: how often do people use terms in a way that exhibits the relevant kind of dual-character behavior? Are they doing so in contexts that we would consider normal, or not? How easy is it to get them to switch their behavior in new contexts? With multiple dimensions of variation on the table, we might propose a number of different thresholds, tied to different dimensions of variation. The question then is this: given the multi-dimensional, gradable nature of the linguistic phenomena at play here, can we draw a threshold in one of these dimensions (or some combination of multiple thresholds, involving different dimensions) whereby positing that some concepts are "dual-character" ones marks a theoretically significant cut, which can in turn do important explanatory and predictive work about our thought and talk? The answer might be "yes". But, absent further discussion, it is hard to see how that line of thinking would best be developed.

These issues also touch on foundational questions about the nature of concepts, and what work (if any) they are meant to play in our thought and talk.[45] What one thinks about these questions will matter for what one thinks it would take to be a special "type" of concept (in a theoretically informative sense, relevant to our discussion here). In turn, that connects to the issue of whether passing a certain threshold tied to linguistic behavior (of the kind we were just discussing) will

suffice to make a concept a special "type" of concept or not, and of what that threshold should be. As we noted at the start of this paper, Knobe, Prasada, and Newman don't say much in their original paper about what exactly concepts in general are, or about what (in general) it would take for there to be a distinctive "kind" of concept (e.g., in the way they claim dual-character concepts are, or natural kind concepts are). The same is true of many other proponents of dual-character concepts.[46] It might well be that turning to these foundational issues will help them provide a principled response to our skepticism about dual-character concepts. We hope that those sympathetic to dual-character concepts take up this line of thinking and see where it goes. We are skeptical that it can be made to work in a fully convincing way, but if nothing else, it might well help clarify the terms of the debate about the existence of dual-character concepts, and whether (and to what extent) the debate is connected to more general disagreements about what "concepts" are, and what role(s), if any, they play in thought and talk.

At this juncture, we want to discuss an important potential fallback option for proponents of dual-character concepts. This is to retreat entirely from the idea that "dual-character concepts" are a unified, theoretically interesting kind of concept, which are marked by having a distinctive internal structure. Instead, theorists interested in the dual-character behavior of how we deploy terms might simply insist that there is a distinctive pattern here in terms of how we deploy certain terms as opposed to others. This move would depart in a substantial way from the original argument from Knobe, Prasada, and Newman, as well as the core of the discussion of dual-character concepts. However, one reason to take it seriously is that, in recent work, a number of theorists interested in dual-character behavior, put more emphasis on the idea of dual-character patterns in how we use language, rather than the idea of dual-character concepts per se. For example, Knobe, Tobia, and Newman, in recent work that significantly complicates the earlier proposal about the relationship between natural kind concepts and dual-character concepts, emphasize the (purportedly distinctive) dual-character patterns involved in our deployment of certain "natural kind" terms (e.g., 'water'). In that work, they leave open a number of distinct possibilities about the implications of such patterns for which theory of natural kind concepts is correct.[47] Given this change in emphasis, it may well be that this sort of "retreat" option is one that some theorists working on (purported) dual-character phenomena would be sympathetic to.

We think this option is worth exploring and might well be an option that proponents of the idea of dual-character phenomena in thought and talk should end up endorsing. One obvious benefit of this option is that it would avoid some of our criticisms of the idea of positing a new, purportedly distinctive kind of concept. That's a significant mark in its favor. But notice that, in light of our studies, it's not even clear what the relevant, purportedly distinctive "dual-character" behavior regarding how we use certain words would amount to. This is because, as we have argued, when we look at the full array of linguistic data that our studies bring out, we don't really see any natural cut-off point where people are using some words but not others in a "dual-character" kind of way. Instead, as we have emphasized, we see a much more varied picture, with lots of flexibility across contexts for when speaker use terms in a way that involves dual-character behavior or not. To reemphasize a point made above: one still needs to draw some kind of threshold within the natural variation in how we use terms in a seemingly "dual-character" way, whether one wants to use that threshold for marking out a new kind of concept or not. As we just argued, given the multi-dimensional, context-dependent, and gradable nature of the relevant types of variation here, it's hard to see what principled cut could be made that would clearly distinguish between those terms that exhibit "dual-character" behavior and those that do not.[48]

5.3 | Appeal to judgments about "true xs" vs. judgments about "good xs"?

Finally, we want to turn to a response tied to an issue we briefly raised (and then put aside) earlier in this paper, in our discussion of Study 2. In their original 2013 paper, Knobe, Prasada, and Neman argue that one good piece of evidence for whether a given concept X has a "dual-character" structure is by asking whether one can felicitously say that something (or someone) "is a true X". Here's how this is meant to meant to work. Recall that, on their account, dual-character concepts "involve two ways of characterizing their instances, and thus two ways of determining category membership,"[49] where the first way concerns an "abstract value" and the second concerns "concrete features" that (generally), when instantiated, are ways of realizing that value. On their account, the modifier 'true' can serve to get people to use the relevant abstract value, rather than the relevant concrete features, for determining whether to apply the concept or not. Thus, they argue that (1) the modifier 'true' can be felicitously applied when using purported dual-character concepts (e.g., "a true scientist"), but (2) cannot be felicitously applied when using purported non-dual-character concepts (e.g., "a true table of contents"), since they do not have the relevant distinctive structure involving some kind of "abstract value". Importantly, the normative criteria involved in something (or someone) being a "true X" are meant to be separate from other normative criteria that may be targeted by statements about something (or someone) being "a good X" (e.g., "a good table of contents"). Given this distinction, one might naturally wonder: can a proponent of dual-character concepts draw on judgments of whether an entity is a "true X", and the distinction from whether an entity is a "good X", to defend the existence of dual-character concepts? We are doubtful that they can.

First, as others have previously pointed out, there are methodological reasons that weigh against drawing on the felicity of "true X" judgments to help support the existence of dual-character concepts. In short, we have good independent reason to think that people in ordinary discourse often use 'true' as a kind of "hedge" (e.g., to signal that one thinks that something counts as a member of a kind, but lacks certain "prototypical" features, or is some kind of edge case) or as an "intensifier" to raise the standards of application for a term. Indeed, it is partly for these reasons that, in recent years, many proponents of dual-character concepts (such as Reuter) have ceased to put much emphasis on the "true-modifier" approach that Knobe, Prasada, and Newman originally emphasized.[50] This is a central reason we decided not to use this approach in our own studies we conducted.

Second, consider again our third study. There, we just took for granted Knobe, Prasada, and Newman's own views about which specific abstract values mattered for the specific terms used. Thus, in this study, we targeted the values they believe to be tied to "true X" judgments, but not "good X" judgments. So, the studies we have relied on in this paper to put pressure on the existence of dual-character concepts already incorporates their ideas about cases involving the specific "abstract values" that are supposed to be relevant to the terms at issue. Moreover, suppose that a proponent of dual-character concepts thinks that Knobe, Prasada, and Newman are wrong about which abstract values are tied to "true X" judgments in the cases involving purported dual-character concepts. Even if so, it's not clear how any other way of specifying these relevant values is going to be able to account for the continuity we've observed between the use of terms that are purportedly tied to dual-character concepts and those that are not.[51]

Finally, suppose we put aside for now the methodological concerns we raised above about the "true-modifier" approach. Now consider again our first study. Recall that, in this study, we focused on when speakers would endorse member and non-member statements using terms tied to

purported non-dual-character concepts. Knobe et al. argue that a certain kind of linguistic behavior about such statements (involving moderate agreement with both) arises from the purportedly distinctive structure of dual-character concepts. But, in our study, we found that such behavior is easily exhibited by terms that are not meant to be linked to such concepts. Given this, we suspect that other forms of linguistic behavior that Knobe et al. associate with dual-character concepts may also be exhibited in such cases. Specifically, we suspect that people would also be willing endorse the felicity of statements like "this is not a true table of contents" in the contexts used in Study 1, or at least do so enough of the time to create trouble for proponents of dual-character concepts. By this, we mean that we suspect they would be willing to do so enough of the time to yield an overall pattern of linguistic data that casts doubt on the idea of there being a sharp line separating those terms tied to purported dual-character concepts and those that are not. Whether that turns out to be so is obviously an empirical question, which deserves further study for anyone who wants to put weight on the "true-modifier" approach.

The combination of these reasons above suggests that it is unlikely that the proponent of dual-character concepts is likely to find much help by appealing to the kinds of "true X" judgments that Knobe et al. emphasized in their original paper. To be clear: this doesn't mean there aren't interesting questions (which might be theoretically important in their own right) to be asked about such judgments, and how, when, and why they come apart from "good X" judgments. If such research unfolds in the right ways, and if the methodological concerns we discussed above can be satisfactorily addressed, perhaps it could be that the proponent of dual-character concepts could draw on this work to support the existence of dual-character concepts. But, given where things stand currently, we don't see much cause for optimism.

5.4 | Summing up

In this section, we considered three different responses on behalf of proponents of dual-character concepts. We don't think that any of these responses is adequate to support the overall case for the existence of dual-character concepts, given our arguments in this paper. Nonetheless, each response raised important theoretical issues and also potential lines for further research. We hope that proponents of dual-character concepts drawn to one (or more) of the above responses take our remarks here as an invitation to further develop these responses, or to delve further into the theoretical issues and potential lines of future research we've discussed.

6 | CONCLUSION

So are there really any dual-character concepts? What we've shown is that the kind of linguistic data that their existence was meant to explain is plausibly better explained by more general accounts of language use. Thus, there are clear reasons for thinking that the relevant linguistic patterns, once fully understood, aren't good evidence for the existence of this kind of concept. Moreover, given the explanatory role that dual-character concepts are meant to play, we've argued that, if they exist, we should expect to see linguistic patterns that we clearly don't see. Thus, by looking more carefully at the linguistic evidence, we see not only a lack of good evidence in favor of positing dual-character concepts, but also some evidence against so doing.[52] Given all of that, we don't think there is good reason to think that there are dual-character concepts, and at least some reason to not posit their existence.

Of course, that doesn't mean that we have given anything close to *definitive* evidence against the existence of this class of concepts. Such a class of concepts may very well exist. After all, it might well be that there are other sources of evidence in favor of positing such concepts, which should lead us to posit their existence, all-things-considered. We hope that proponents of dual-character concepts can see our paper as an invitation to say more about what those other sources of evidence might be. Yet, as of now, we don't see any strong evidence on the table that favors believing in dual-character concepts, all-things-considered. Thus, given the evidence we have put on the table against so doing, we think there is a strong initial case for not believing in them.

In closing, we want to be clear that our main lines of argument in this paper are not motivated by some general skepticism about the usefulness or existence of concepts.[53] Similarly, we are not against making concrete proposals about the structure of cognition and the often-important role of normative evaluations in such structure. Much of the work of one of the authors of this paper (Phillips) makes precisely these sorts of proposals, including in co-authored work with Knobe.[54] And finally, we want to be clear that we are not generally skeptical about the idea that patterns of linguistic data can be informative about non-linguistic cognition. For example, in recent joint work with Angelika Kratzer, one of us (Phillips) has offered an account of how to fit together the semantics for modal language with the non-linguistic cognition that allows for the representation of possibilities.[55] This account draws heavily on showing parallels between modal language and non-linguistic tasks that involve reasoning about possibilities and argues that the two are mutually informative. So, we're not generally skeptical about there being deep and important intersections between empirical work in cognitive science and work in semantics, philosophy of language, or philosophy of mind. In fact, in the case of dual-character concepts, we think our conclusions suggest that more work in this vein needs to be done, not less.

ACKNOWLEDGMENTS

We would like to thank Nat Hansen, Justin Khoo, Sean Kim, Josh Knobe, Zachary Lang, Daniel Lassiter, Shen-Yi Liao, Tristram McPherson, Eliot Michaelson, Adrian Russian, and Rachel Sterken for helpful discussions and feedback. And we especially want to thank Zoe Thierfelder for her help in creating and conducting the studies, and for her insights throughout our work on this project.

ENDNOTES

[1] (Knobe, Prasada, and Newman 2013).

[2] (Knobe, Prasada, and Newman 2013, 243).

[3] (Knobe, Prasada, and Newman 2013, 243).

[4] (Knobe, Prasada, and Newman 2013, 243).

[5] In this paper, we use small caps (e.g., PIANO) to pick out concepts, single quotation marks (e.g., 'piano') to mention linguistic items, and double quotation marks (e.g., "piano") for a variety of tasks, including quoting others' words, scare quotes, and mixes of use and mention.

[6] (Knobe, Prasada, and Newman 2013, 243).

[7] (Knobe, Prasada, and Newman 2013, 243).

[8] On this front, the idea of dual-character concepts can be usefully contrasted with Kevin Scharp's work on "defective concepts", which he thinks (put roughly) involve application-conditions that conflict in certain cases, and thus call both for applying the concept and not. See (Scharp 2020).

[9] See, for example, (Knobe 2022) on personal identity, (Liao, Meskin, and Knobe 2020) on art, (Flanagan and Hannikainen 2022) on law, (Guo, Dweck, and Markman 2021) on gender, and (Newman and Knobe 2019) on natural kind concepts.

[10] (Leslie 2015).

[11] (Haslanger 2020).

[12] (Tobia, Newman, and Knobe 2020).

[13] See (Leslie 2015) and (Del Pinal and Reuter 2017). See (Reuter 2019) for an overview and further citations.

[14] (Leslie 2015).

[15] For an example of this kind of view, see (Gibbard 2003).

[16] It should be noted that (as we briefly discuss later in this paper), in recent work, two of the co-authors of this original paper (Knobe and Newman) argue for a more complicated picture of the relationship between natural kind concepts and dual-character concepts, on which natural kind concepts also have a certain kind of "dual-character" structure. See (Newman and Knobe 2019) and (Tobia, Newman, and Knobe 2020).

[17] These assumptions are obviously controversial in and of themselves, but we take them on here to make it easier to straightforwardly engage with the argument put forward by (Knobe, Prasada, and Newman 2013). We return to some of the relevant general issues tied to the controversies here (including ones about what work "concepts" are meant to do in explaining thought and talk) after presenting our studies.

[18] We tested the full list of 21 terms thought to be associated with non-dual-character concepts. These were: 'cashier', 'catalog', 'second cousin', 'chair', 'obituary', 'uncle', 'stroller', 'welder', 'rustling', 'firefighter', 'blog', 'caseworker', 'mechanic', 'optician', 'baker', 'doorman', 'mayor', 'waitress', 'tailor', 'bartender', and 'table of contents'.

[19] As in (Knobe, Prasada, and Newman 2013), participants also completed two "catch" questions which included a case of an unambiguously clear instance of a person who satisfies the criteria typically associated with the concept BIOLOGIST (a woman who was one of the world's best biologists and who does nothing but biology), and a case of an unambiguously clear instance of an animal that failed to satisfy the criteria that are typically associated with the concept DOG (a cat who was not in any way dog-like). After reading these cases, participants rated their agreement with the member and non-member assertions involving the terms 'biologist' and 'dog' (respectively). Participants were excluded if their responses did not fall within the correct quarter of the agreement scale for all four assertions, leading us to exclude the data from 21 participants.

[20] Here we use simple bar graph depictions of the data here to facilitate comparisons to the patterns as they were originally reported by Knobe, Prasada, and Newman (see Fig. 1b). This way of depicting data can obscure underlying variation, and we return to precisely this variation in §3.1.

[21] We analyzed the data using a comparison of linear mixed-effects models which allow us to model the variation in these effects across both different concepts and different participants. For those interested, all of our materials, data, analysis code, and a technical summary of the experimental and analysis details are available here: https://doi.org/10.7910/DVN/DEHRG6

[22] See, for example, Leslie's discussion of such a case involving 'bus driver' in (Leslie 2015, 122-123) and her discussion of how one can create contexts to "coerce" (Leslie 2015, 125) the term 'bartender' to exhibit dual-character behavior.

[23] We tested the full list of 20 terms Knobe et al. thought to be associated with dual-character concepts. These were: 'artist', 'criminal', 'love', 'minister', 'mentor', 'comedian', 'theory', 'boyfriend', 'friend', 'argument', 'teacher', 'art museum', 'poem', 'soldier', 'sculpture', 'musician', 'mother', 'novel', 'scientist', and 'rock music'.

[24] As in the prior study, participants completed two "catch" questions, and we again applied the same exclusion criteria, resulting in the exclusion of data from 19 participants.

[25] These statistical tests were one-sample t-tests with an alternative of 50 (the midpoint of the agreement scale) indicating that participants gave ratings that were significantly above the midpoint.

[26] We analyzed the data using a comparison of linear mixed-effects models which allow us to model the variation in these effects across both different concepts and different participants. All of our materials, data, analysis code, and a technical summary of the experimental and analytic details are available here: https://doi.org/10.7910/DVN/DEHRG6

[27] Thanks to Josh Knobe for this suggestion and generous discussion.

[28] In more recent work on dual-character linguistic behavior, (Tobia, Newman, and Knobe 2020) suggest that this behavior (at least in the case of natural kind concepts) may be sensitive to which of the two criteria are emphasized in the context. Here, we pursue a related question, but use the original proposed set of dual-character concepts from (Knobe, Prasada, and Newman 2013), and return to their more recent proposal from (Tobia, Newman, and Knobe 2020) later.

[29] In this case, we tested the 10 terms argued to be associated with dual-character concepts: 'artist', 'criminal', 'love', 'minister', 'mentor', 'comedian', 'theory', 'boyfriend', 'friend', and 'argument'.

[30] A fourth study that we conducted—reported in the supplement but not included in full in the main text of this paper, for the sake of brevity—verified this hypothesis. The study was very similar to Study 3 except for two important changes. The first was the difference in the formulation of the member and non-member assertions, which resembled those used in Study 2. The second was that we had the description of the entity, which highlighted the entity's failure to satisfy some normative criteria, occur after the description of the conversational context. Both changes served to make the lacking normative criteria more relevant in the conversational context. In line with our hypothesis above, we found that the agreement patterns in both conversational contexts were more similar to one another in this case. A full description of the study can be found in the repository here: https://doi.org/10.7910/DVN/DEHRG6

[31] As in the prior two studies, participants also completed two "catch" questions and we again applied the same exclusion criteria, resulting in the exclusion of data from 18 participants.

[32] (Knobe, Prasada, and Newman 2013, 250).

[33] We analyzed the data using a comparison of linear mixed-effects models which allow us to model the variation in these effects across both different concepts and different participants. All of our materials, data, analysis code, and a technical summary of the experimental and analytic details are available here: https://doi.org/10.7910/DVN/DEHRG6

[34] (Knobe, Prasada, and Newman 2013, 250).

[35] It is worth pointing out that the size of this difference is smaller in this study than in the previous studies. One possible explanation is that the larger differences found in the prior studies were a result of the additional modifiers used in the member and non-member statements. Another possibility is that the smaller effect was observed in this study because the manipulation of the conversational context was comparatively subtle (or perhaps was in tension with the thrust of the conversational context created in the original descriptions created by Knobe, Prasada, and Newman). We return to questions about this kind of variation in §3.1.

[36] See (Plunkett and Sundell 2013).

[37] For other views close to Plunkett and Sundell's account in key respects, see (Thomasson 2016) and (Khoo 2020). For views that involve some similar ideas, but push for more radical views about how pervasive shifts in meaning are, see (Ludlow 2014) and (Stolk, Bašnáková, and Toni 2022).

[38] See (Plunkett, Sterken, and Sundell 2023), which argues for a view that combines the generical account of generics put forward in (Sterken 2015) with the view of metalinguistic negotiation put forward in (Plunkett and Sundell 2013). For Leslie on normative generics, see (Leslie 2015).

[39] See (Tobia, Newman, and Knobe 2020).

[40] (Kiel 1989).

[41] In Study 3, participants only rated their agreement with either the member or the non-member statement, so we don't have enough data to determine whether each participants' response to each term exhibits dual-character behavior. However, the group-level distributions tell much the same story.

[42] (Lakoff 1973) and (Malt 1990).

[43] (Gellman 2003).

[44] (Knobe, Prasada, and Newman 2013, 245).

[45] For some different takes here, see (Margolis and Laurence 1999).

[46] For example, see how (Reuter 2019) introduces and discusses the basic topic of dual-character concepts in his overview of recent work on the topic.

[47] (Tobia, Newman, and Knobe 2020).

[48] This same basic point carries over to other potential "retreat" options from the idea of dual-character concepts (understood as something about cognition) to other claims solely at the level of language. For example, consider Leslie's view that "dual character concepts give rise to lexical entries that have distinct, though related, senses" (Leslie 2015, 116). On her view, a word (e.g., 'scientist') used to express a dual-character concept has two different meanings, one of which is "normative" (tied to the relevant "abstract value") and the other is "descriptive" (tied to the relevant concrete features). One might endorse Leslie's claim about polysemy regarding the relevant terms without also (as she does) endorsing the idea of dual-character concepts as such (which we take her to agree is a view at the level of cognition). However, for the reasons given above, even if it is true that the relevant terms are polysemous, it's hard to see what warrants positing that *these* polysemous terms (and not others) pick out a distinctive, theoretically interesting class of terms worth discussing as "dual-character" ones. After all, the (purported) reason for doing that is that these terms (and not others)

exhibit a distinctive kind of dual-character behavior. Our data about variation, however, challenges that claim.

[49] (Knobe, Prasada, and Newman 2013, 243).

[50] See (Reuter 2019, 3).

[51] It is worth noting that proponents of dual-character concepts who have put different views about how the relevant "abstract values" are related to the relevant "concrete features" (e.g., Leslie and Reuter) haven't disputed the basic extensional claims made by Knobe et al. on which specific abstract values matter in the core cases of dual-character concepts that they studied. Thus, while it might be possible to insist that there are dual-character concepts but that Knobe et al. have just been wrong about which abstract values matter in the cases at hand, this would involve a significant departure from the main lines of existing discussion about dual-character concepts. It should be noted that if a proponent of dual-character concepts were to propose other abstract values here for the (purported) dual-character concepts at hand, it would then make sense to replicate the third study using whatever abstract values are proposed. We invite proponents of such a view to pursue precisely this work.

[52] This inference assumes Bayesian absent evidence reasoning. See (Stephens 2011) and (Hsu et al. 2017).

[53] For an example of a skeptical take on concepts, see (Machery 2009).

[54] See, for example, (Phillips and Knobe 2018), (Phillips and Cushman 2017), and (Phillips et al. 2017).

[55] (Phillips and Kratzer Manuscript), drawing on (Kratzer 2012).

REFERENCES

Del Pinal, Guillermo, and Kevin Reuter. 2017. Dual Character Concepts in Social Cognition: Commitments and the Normative Dimension of Conceptual Representation. *Cognitive Science* 41 (S3):477–501.

Flanagan, Brian, and Ivar R. Hannikainen. 2022. The Folk Concept of Law: Law Is Intrinsically Moral. *Australasian Journal of Philosophy* 100 (1):165-179.

Gellman, Susan. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.

Guo, Cai, Carol S. Dweck, and Ellen M. Markman. 2021. Gender Categories as Dual-Character Concepts? *Cognitive Science* 45 (5):e12954.

Haslanger, Sally. 2020. Going On, Not in the Same Way. In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen and D. Plunkett. Oxford: Oxford University Press. 230–260.

Hsu, Anne S, Andy Horng, Thomas L Griffiths, and Nick Chater. 2017. When Absence of Evidence is Evidence of Absence: Rational Inferences from Absent Data. *Cognitive science* 41:1155-1167.

Khoo, Justin. 2020. Quasi Indexicals. *Philosophy and Phenomenological Research* 100 (1):26-53.

Kiel, Frank C. 1989. *Concepts, Kinds, and Conceptual Development*. Cambridge, MA: MIT.

Knobe, Joshua. 2022. Personal Identity and Dual Character Concepts. In *Experimental Philosophy of Identity and the Self*, edited by K. Tobia: Bloomsbury.

Knobe, Joshua, Sandeep Prasada, and George E. Newman. 2013. Dual Character Concepts and the Normative Dimension of Conceptual Representation. *Cognition* 127 (2):242-257.

Kratzer, Angelika. 2012. *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press.

Lakoff, George. 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2 (4):458–508.

Leslie, Sarah-Jane. 2015. "Hillary Clinton is the Only Man in the Obama Administration": Dual Character Concepts, Generics, and Gender. *Analytic Philosophy* 56 (2):111-141.

Liao, Shen-yi, Aaron Meskin, and Joshua Knobe. 2020. Dual Character Art Concepts. *Pacific Philosophical Quarterly* 101 (1):102-128.

Ludlow, Peter. 2014. *Living Words: Meaning Undetermination and the Dynamic Lexicon*. Oxford: Oxford University Press.

Machery, Edouard. 2009. *Doing Without Concepts*: Oxford University Press.

Malt, Barbara C. 1990. Features and Beliefs in the Mental Representation of Categories. *Journal of Memory and Language* 29:289-315.

Margolis, Eric, and Stephen Laurence. 1999. *Concepts: Core Readings*. Cambridge: MIT Press.

Newman, George E., and Joshua Knobe. 2019. The Essence of Essentialism. *Mind and Language* 34 (5):585-605.

Phillips, Jonathan, and Angelika Kratzer. Decomposing Modal Thought. Unpublished manuscript, November 2023.

Phillips, Jonathan, and Fiery Cushman. 2017. Morality Constrains the Default Representation of What is Possible. *Proceedings of the National Academy of Sciences of the United States of America* 114 (18):4649-4654.

Phillips, Jonathan, and Joshua Knobe. 2018. The Psychological Representation of Modality. *Mind and Language* 33 (1):65-94.

Phillips, Jonathan, Christian Mott, Julian De Freitas, June Gruber, and Joshua Knobe. 2017. True Happiness: The Role of Morality in the Folk Concept of Happiness. *Journal of Experimental Psychology: General* 146 (2):165-181.

Plunkett, David, Rachel Katharine Sterken, and Timothy Sundell. 2023. Generics and Metalinguistic Negotiation. *Synthese* 201 (50):1-46.

Plunkett, David, and Tim Sundell. 2013. Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint* 13 (23):1-37.

Reuter, Kevin. 2019. Dual Character Concepts. *Philosophy Compass* 14 (1):e12557.

Scharp, Kevin. 2020. Philosophy as the Study of Defective Concepts. In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen and D. Plunkett. Oxford: Oxford University Press. 396-416.

Stephens, Christopher Lee. 2011. A Bayesian Approach to Absent Evidence Reasoning. *Informal Logic* 31 (1):56-65.

Sterken, Rachel Katharine. 2015. Generics in Context. *Philosophers' Imprint* 15.

Stolk, Arjen, Jana Bašnáková, and Ivan Toni. 2022. Joint Epistemic Engineering: The Neglected Process in Human Communication. In *The Routledge Handbook of Semiosis and the Brain*, edited by A. M. García and A. Ibáñez. New York: Routledge. 259–278.

Thomasson, Amie L. 2016. Metaphysical Disputes and Metalinguistic Negotiation. *Analytic Philosophy* 57 (4):1-28.

Tobia, Kevin P., George E. Newman, and Joshua Knobe. 2020. Water is and is Not H 2 O. *Mind and Language* 35 (2):183-208.

How to cite this article: Phillips, J. & Plunkett, D. (2023). Are there really any dual-character concepts?. *Philosophical Perspectives*, 37: 340–369. https://doi.org/10.1111/phpe.12194