Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
doi: 10.1111/phpr.12217

Aesthetic Adjectives: Experimental Semantics and Context-Sensitivity

SHEN-YI LIAO
University of Leeds / University of Puget Sound

AARON MESKIN
University of Leeds

One aim of this essay is to contribute to understanding *aesthetic communication*—the process by which agents aim to convey thoughts and transmit knowledge about aesthetic matters to others. Our focus will be on the use of aesthetic adjectives in aesthetic communication. Although theorists working on the semantics of adjectives have developed sophisticated theories about gradable adjectives, they have tended to avoid studying aesthetic adjectives—the class of adjectives that play a central role in expressing aesthetic evaluations (e.g., 'beautiful', 'ugly', 'elegant'). And despite the wealth of attention paid to aesthetic adjectives by philosophical aestheticians, they have paid little attention to contemporary linguistic theories of adjectives. We take our work to be a first step in remedying these lacunae. In this paper, we present four experiments that examine one aspect of how aesthetic adjectives ordinarily function: the context-sensitivity of their application standards. Our results present a *prima facie* empirical challenge to a common distinction between *relative* and *absolute* gradable adjectives because aesthetic adjectives are found to behave differently from both. Our results thus also constitute a *prima facie* vindication of some philosophical aestheticians' contention that aesthetic adjectives constitute a particularly interesting segment of natural language, even if the boundaries of this segment might turn out to be different from what they had in mind.

Philosophical aestheticians are interested in a wide range of things having to do with the aesthetic domain, such as the production of art, the reception of art, our aesthetic responses to nature, and the nature of art itself. But communication is another key aspect of our aesthetic lives with which aestheticians have been and should be concerned. After all, we don't just produce and consume art—we spend a significant amount of time talking, writing, and reading about it. One aim of this essay is to contribute to understanding *aesthetic communication*—the process by which agents aim to convey thoughts and transmit knowledge about aesthetic matters to

others.[1] Our focus will be on the use of aesthetic adjectives in aesthetic communication, and in focusing on such adjectives we follow in a long line of contemporary aestheticians who have been inspired by Frank Sibley's (1959, 1965) seminal work on aesthetic concepts.

So how do aesthetic adjectives work? Although theorists working on the semantics of adjectives have developed sophisticated theories about gradable adjectives generally and have explored the closely related class of adjectives known as predicates of personal taste (e.g., 'tasty', 'fun'), they have tended to avoid studying aesthetic adjectives—the class of adjectives that play a central role in expressing aesthetic evaluations (e.g., 'beautiful', 'ugly', 'elegant'). And despite the wealth of attention paid to aesthetic adjectives by philosophical aestheticians, they have paid little attention to contemporary linguistic theories of adjectives. We take our work to be a first step in remedying these lacunae.

In this paper, we present four experiments that examine one aspect of how aesthetic adjectives ordinarily function: the context-sensitivity of their application standards. Our results present a *prima facie* empirical challenge to a common distinction between *relative* and *absolute* gradable adjectives because aesthetic adjectives are found to behave differently from paradigmatic relative and absolute gradable adjectives. Our results thus also constitute a *prima facie* vindication of some philosophical aestheticians' (e.g., Sibley 1959, 2001a, 2001b) contention that aesthetic adjectives constitute a particularly interesting segment of natural language, even if the boundaries of this segment might turn out to be different from what they had in mind.

Section 1 briefly presents the philosophical and psycholinguistic background for our experiments. Sections 2–5 report the methods and results of our four experiments. Section 6 presents a dilemma for the distinction between relative and absolute gradable adjectives and discusses ways that one might escape the dilemma. Section 7 discusses further implication of our results—in relation to predicates of personal taste, experimental philosophy, and aesthetic communication—and outlines avenues for future research.

1. Classifying Gradable Adjectives

1.1. Relative vs. Absolute

The majority, if not all, of evaluative aesthetic terms (e.g., 'beautiful', 'ugly') are *gradable adjectives*. Like other gradable adjectives, they admit of comparative constructions, such as 'Barbara Hepworth's sculptures are *more* beautiful than Henry Moore's'.

[1] So our conception of aesthetic communication is quite distinct from that of Gary Iseminger (2004: 25) who understands 'aesthetic communication' to refer to the process by which an agent designs and produces an artifact with the aim and result that it is appreciated by some other agent.

One common classification scheme of gradable adjectives concerns their context-sensitivity. Some gradable adjectives, such as 'long' and 'tall', are typically interpreted relative to a contextually-determined comparison class. Other gradable adjectives, such as 'spotted' and 'flat', typically are not. Following linguists Christopher Kennedy and Louise McNally (2005), we will call the former kind *relative gradable adjectives* (or 'relative' for short) and the latter kind *absolute gradable adjectives* (or 'absolute' for short). And we will call the classification scheme *the relative/absolute classification scheme of gradable adjectives*.

Absolute adjectives come in two varieties—*minimal standard* adjectives (e.g., 'damp', 'bent'), which require some minimal degree of the relevant property, and *maximal standard* adjectives (e.g., 'full', 'closed'), which require a maximal degree of the relevant property. Although such adjectives may be used imprecisely in certain contexts (e.g., one might call a shopping bag full even if there is a little room left in it), they do not admit of context-dependent interpretations.[2]

1.2. Non-Experimental Evidence

One piece of semantic data for the relative/absolute classification scheme of gradable adjectives concerns the different patterns of entailment that relative and absolute adjectives generate.[3] Consider first an example with a relative adjective. The statement 'John is taller than Jim' entails neither 'John is tall' nor 'Jim is not tall'. Consider second an example with a maximal standard adjective. The statement 'his fence is straighter than your fence' seems to entail 'your fence is not straight'. Similarly, consider a minimal standard absolute adjective. The statement 'her fence is more bent than our fence' seems to entail 'her fence is bent'. Typically, comparative statements involving absolute adjectives generate such entailments, but comparative statements involving relative adjectives do not (Kennedy 2007).

At first glance, many positive aesthetic adjectives appear to be best classified as relative adjectives. 'John is more beautiful than Jim' entails neither 'Jim is not beautiful' nor 'John is beautiful'. 'Jane is prettier than Jan' entails neither 'Jane is pretty' nor 'Jan is not pretty'.[4] However, discussions of aesthetic adjectives are largely absent from the inquiry into the nature of

---

[2] See Kennedy (2007: 23–25) for discussion of imprecision.

[3] Toledo and Sassoon (2011) and Liao, McNally, and Meskin (forthcoming) survey other linguistic diagnostics in support of, and against, the relative/absolute classification scheme of gradable adjectives.

[4] Bierwisch (1989: 206–207) claims that it follow from 'Eva is prettier than Helga' that both Eva and Helga are pretty. If so, 'pretty' may be a minimal standard absolute adjective. We are unconvinced: 'Eva is not pretty but she is prettier than Helga' sounds fine to us. The equivocal nature of these semantic intuitions may play a role in explaining the results of our studies.

gradable adjectives. We will argue that overlooking aesthetic adjectives is a mistake because aesthetic adjectives turn out to present puzzling data that complicate the relative/absolute classification scheme of gradable adjectives.

1.3. Experimental Evidence

Building on the relative/absolute classification scheme, linguist Kristen Syrett and colleagues have developed an experimental paradigm, the *presupposition assessment task (PAT)*, for classifying gradable adjectives as either relative or absolute depending on their patterns of use by competent speakers (Syrett *et al.* 2006, 2010). In the main studies, they presented children and adults with pairs of objects with different degrees of the relevant property. They then asked them to pick out *the* long one, *the* spotted one, and so on. According to them, the use of the definite article 'the' in such requests involves two presuppositions: EXISTENCE (that there is at least one object satisfying the adjective) and UNIQUENESS (that there is at most one object satisfying the adjective).

On a standard view, a key feature of relative adjectives is that they have *standards of comparison* that vary by context. For example, the cut-off point for counting an object as being long varies according to the salient objects of comparison in a given context. Hence, when relative adjectives are used in the request, people are typically able to construct a comparison class "on the fly"—namely, by shifting the standard of comparison—so that both EXISTENCE and UNIQUENESS are satisfied. Experimentally, Syrett and colleagues found that both children and adults did typically *comply* with the request in the PAT when it came to paradigmatic relative adjectives, such as 'big' and 'long'. For example, when presented with two rods of different lengths, both children and adults typically picked out the longer rod as *the* long one. Participant compliance indicates that they are able to construct a comparison class so that both EXISTENCE and UNIQUENESS are satisfied, which indicates that the adjective in question is best classified as relative.

In contrast with relative adjectives, absolute adjectives have context-independent standards of comparison. Hence, requests involving absolute adjectives will not always allow for the construction of a comparison class so that both EXISTENCE and UNIQUENESS are satisfied. For example, regardless of context, an object counts as being bent if it is bent to a non-zero degree. Faced with two rods bent to different (non-zero) degrees, people are not typically able to construct a comparison class "on the fly" in which one counts as bent and the other does not. So in that case UNIQUENESS cannot be satisfied. Experimentally, Syrett and colleagues found that adults (and children, with some exceptions described below) did typically *refuse* the request in the PAT when it came to paradigmatic absolute adjectives, such as 'spotted', 'full', 'bumpy' and 'straight'. For example, when presented with two

discs which are both spotted to different degrees, both children and adults typically refuse to pick out the spotted one since UNIQUENESS is violated. Most adults also refuse when asked to pick out the full one between two jars neither of which are full but are full to different degrees. Again, this is due to the violation of UNIQUENESS. And although children exhibit a different pattern of behavior with 'full' and 'straight' (they are significantly more willing than adults to choose the more full one in response to the request for the 'full' one and the straighter one in response to the request for the 'straight' one), Syrett and her colleagues argue, on the basis of reaction time studies, that they still exhibit a marked difference in their treatment of relative and absolute adjectives. For example, even when they picked out the more full jar as the full one, children typically took much longer to do so. Participant refusal, or a notable increase in reaction time, indicates that the adjective in question does not have a context-dependent standard of comparison, and is thus best classified as absolute.

2. Study 1: 'Beautiful'

2.1. Motivation and Methods

We implemented the PAT as an online questionnaire. Participants were given instructions of the task alongside an instructional manipulation check that ensures participants read the instructions carefully (cf. Oppenheimer *et al.* 2009). Each adjective / stimuli set was then presented in the following format (fig. 1):

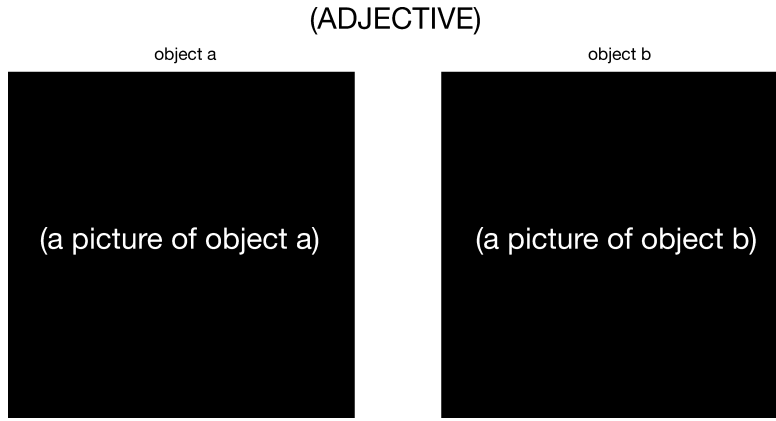

Figure 1. *adjective / stimuli set presentation format for Study 1 & Study 2*

Participants were then asked to choose one response from the following options for each adjective / stimuli set (the order of the options remained constant).

- Object A is the (ADJECTIVE) object.
- Object B is the (ADJECTIVE) object.
- I can't. Neither Object A nor Object B is (ADJECTIVE).
- I can't. Both Object A and Object B are (ADJECTIVE).

For this study, we simply counted the first and second responses as *compliance* responses, and the third and fourth responses as *refusal* responses. All participants went through 8 adjective / stimuli sets (the order of presentation was counterbalanced): disks that are *spotted* to different degrees, rods that are *bent* to different degrees, blocks that are *long* to different degrees, and male faces that are *beautiful* to different degrees.[5]

For each adjective tested there were two comparison sets of stimuli, which were constructed using a combination of existing photographs and digital manipulation. For example, for the 'beautiful' stimuli we modified a photograph of a man's face to produce three versions with increasing feature asymmetry. One comparison set consisted of the least asymmetric one and the intermediate one. Another comparison set consisted of the intermediate one and the most asymmetric one.

We recruited 40 participants (19 women; M_{age} = 34.03; SD_{age} = 14.27) for this study from Amazon Mechanical Turk, restricting the eligibility to people whose registered location is in the United States and have HIT approval rate greater than or equal to 95%.[6]

2.2. Results and Discussion

We replicated Syrett and colleagues' results with non-aesthetic gradable adjectives. 97.4% of participants complied with the request to pick out the long object. In contrast, only 17.9% of participants complied with the request to pick out the straight object (where EXISTENCE is violated) and only 10.3% of participants complied with the request to pick out the spotted object (where UNIQUENESS is violated). We did not find any gender or order effects, and so these variables are not investigated further.

We found that 43.6% of participants complied with the request to pick out the beautiful object amongst two male faces. When we compared patterns of compliance and refusal between the adjectives tested (fig. 2), it

[5] Full material and data for all studies reported in this paper are deposited in the Open Science Framework repository and openly available for access at https://osf.io/6uztd/. The rod and block images are drawn from Syrett and colleagues' stimuli (cf. Syrett 2007). The unmanipulated face images are drawn from an emotion expression database and chosen for their emotion neutrality.

[6] See Paolacci and Chandler (2014) for an overview of the demographic characteristics of the Amazon Mechanical Turk participant pool, its validity for conducting social scientific research, and data quality comparisons with traditional university lab studies.

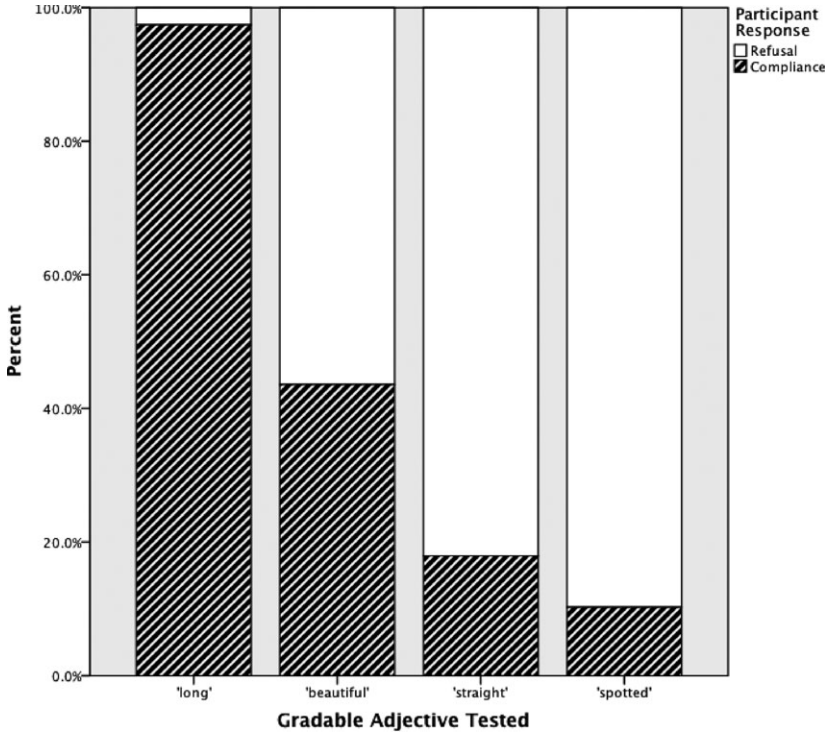

Figure 2. *'beautiful' vs. paradigmatic relative and absolute adjectives*

turns out that 'beautiful' functioned very differently from the relative adjective 'long' ($X^2(1) = 54.384$, $p < 0.001$, Cramer's $V = 0.590$), and also from the absolute adjectives 'straight' ($X^2(1) = 7.510$, $p = 0.006$, Cramer's $V = 0.253$) and 'spotted' ($X^2(1) = 13.173$, $p < .001$, Cramer's $V = 0.336$). If 'beautiful' is typical of aesthetic adjectives, then the results from Study 1 suggest that aesthetic adjectives do not function like either paradigmatic relative adjectives or paradigmatic absolute adjectives.

3. Study 2: 'Ugly'

3.1. Motivation and Methods

Although Study 1 is suggestive, it does not clearly establish the general conclusion that aesthetic adjectives function differently from both relative and absolute adjectives. One might wonder whether the results of Study 1 merely reflect experimental artifacts rather than a genuine pattern of language usage. Specifically, first, one might wonder whether the same result holds of other aesthetic adjectives, and, second, one might wonder whether the same result holds of different kinds of stimuli.

We wanted to address these doubts in Study 2. We tested the negatively-valenced aesthetic adjective 'ugly', in part because some semantic diagnostics suggest that it—unlike 'beautiful'—may be better classified as absolute rather than relative.[7] We also used a range of stimuli from different domains: people (digitally manipulated photographs of a female face with different levels of asymmetry), artifacts (photographs of sports cars in different stages of restoration), and natural objects (photographs of sunflowers in different stages of decay). As before, we followed Syrett and colleagues in using 'long' as our paradigmatic relative adjective and 'spotted' as our paradigmatic absolute adjective.

The adjective / stimuli sets were presented in the same format as before (fig. 1). However, for this study, we conducted a further analysis that made use of another phase of the experiment. As was the case in Study 1, there was a phase of the experiment—call it the *selective phase*—where participants were asked to make a *selective judgment* about the stimuli with respect to the salient adjective. In the selective phase, we asked participants to choose one response from the following options for each adjective / stimuli set:

- Object A is the (ADJECTIVE) object.
- Object B is the (ADJECTIVE) object.
- I can't. Neither Object A nor Object B is (ADJECTIVE).
- I can't. Both Object A and Object B are (ADJECTIVE).

After the selective phase, participants entered another phase of the experiment—call it the *comparative phase*—where they were asked to make a *comparative judgment* about the stimuli with respect to the salient adjective. In the comparative phase, we asked participants to choose one response from the following options for each adjective / stimuli set:

- Object A is more (ADJECTIVE) than Object B.
- Object B is more (ADJECTIVE) than Object A.
- Neither object is (ADJECTIVE).
- Both objects are equally (ADJECTIVE).

Within each phase, the order of presentation of the adjective / stimuli sets was randomized and the order of response options remained constant.

---

[7] What sort of gradable adjective is 'ugly'? 'John is uglier than Jim' does not seem to entail 'Jim is not ugly'. So 'ugly' does not seem to be a maximal standard absolute adjective. But Morzycki (2012: 578) suggests that 'ugly' has a scale closed at the lower end. If so, it is a minimal standard absolute adjective.

Results from the comparative phase can provide insights into the *reasons* for participants' responses in the selective phase.[8] Some participants refused to judge one object as more beautiful or ugly than the other. Call such responses *neither-greater responses*. If a participant's refusal response in the selective phase is linked to a neither-greater response in the comparative phase, then the former response would not constitute evidence for her treating an adjective as an absolute adjective. Hence, to ensure that participant response patterns gathered from the selective phase are truly indicative of their treatment of a gradable adjective as relative or absolute, in subsequent analyses we *filtered out* the refusal responses in the selective phase that merely stem from neither-greater responses in the comparative phase.

We recruited 40 participants (11 women; M_{age} = 26.85; SD_{age} = 7.49) for this study from Amazon Mechanical Turk, again restricting the eligibility to people whose registered location is in the United States and have HIT approval rate greater than or equal to 95%.

3.2. Results and Discussion

Although semantic tests indicate that the aesthetic adjective 'ugly' is classifiable as a gradable adjective, a decent proportion of participants in Study 2 (37.1% for people; 12.8% for artifacts; 23.1% for natural objects) were unwilling to judge one object as uglier than another (fig. 3). In contrast, virtually all participants were willing to judge that one object possessed the relevant property to a greater degree in the case of the paradigmatic relative and absolute adjectives. In other words, there is a relative prevalence of neither-greater responses with 'ugly'. We think this phenomenon calls for an explanation in its own right, but we have no settled view yet.

We then filtered out neither-greater responses from subsequent analyses so that we can determine whether the aesthetic adjective 'ugly' is relative or absolute. When we compared patterns of compliance and refusal between the adjectives tested, we observed the same kind of results that we found in Study 1 (fig. 4). With people as stimuli, 'ugly' functioned differently from 'long' (Fisher's exact (2-sided)[9] , p = 0.029, Cramer's V = 0.309) and from

---

[8] Results from the comparative phase also allowed us to exclude *obvious nonsense responses*, such as picking Object A to be the ugly one but judging that Object B is uglier than Object A. On this basis, 5 responses were excluded from 'ugly' (face), 1 response was excluded from 'ugly' (car), 1 response was excluded from 'ugly' (flower), and 2 responses were excluded from 'long'. This exclusion criterion was determined and implemented prior to any substantial analysis of the data.

[9] Fisher's exact test is used wherever the minimum-expected-cell-count assumption of Pearson's chi-square test is violated. There is no clear consensus on the best way to report effect size for Fisher's exact test, but Cramer's V is somewhat accepted, and we report it for ease of comparison with other results.

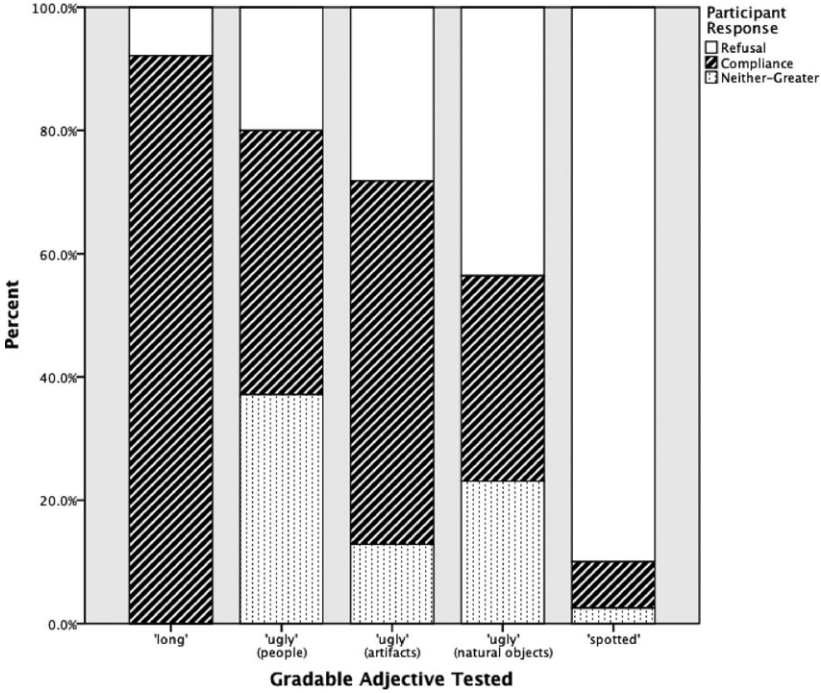

Figure 3. *'ugly' vs. paradigmatic relative and absolute adjectives (refusal/compliance/neither-greater)*

'spotted' ($X^2(1) = 24.742$, $p < 0.001$, Cramer's $V = 0.637$). With artifacts as stimuli, 'ugly' functioned differently from 'long' ($X^2(1) = 6.853$, $p = 0.009$, Cramer's $V = 0.309$) and from 'spotted' ($X^2(1) = 28.474$, $p < 0.001$, Cramer's $V = 0.625$). With natural objects as stimuli, 'ugly' functioned differently from 'long' ($X^2(1) = 19.208$, $p < 0.001$, Cramer's $V = 0.531$) and from 'spotted' ($X^2(1) = 12.093$, $p = 0.001$, Cramer's $V = 0.419$). Collectively, the results of Study 2 again indicate that aesthetic adjectives do not function like either paradigmatic relative adjectives or paradigmatic absolute adjectives.

4. Study 3: Abstract Sculptures

4.1. Motivation and Methods

Although we believe that Study 1 and Study 2 together give fairly strong evidence that aesthetic adjectives complicate the relative/absolute classification scheme of gradable adjectives, further questions arose when we presented the earlier experiments. We group the questions into two kinds.

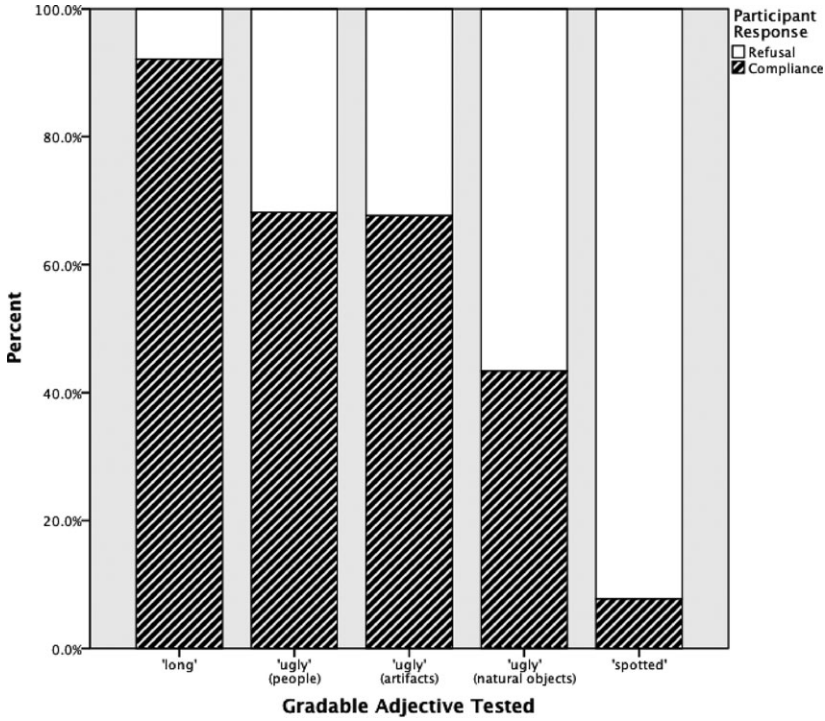

Figure 4. *'ugly' vs. paradigmatic relative and absolute adjectives (refusal/compliance)*

Theoretical:

- One of our investigative goals is to increase philosophical aestheticians' understanding of the language of aesthetics. In the philosophical aesthetics literature, discussions of aesthetic terms have often focused on their use in artistic contexts (Sibley 1959; Walton 1970; Kivy 1973). However, Study 1 and Study 2 only test for the use of aesthetic adjectives with mundane objects, such as people and artifacts. So, do ordinary people apply aesthetic adjectives to artworks in the same way that they apply aesthetic adjectives to mundane objects?
- One limitation with generalizing from 'beautiful' and 'ugly' is that they are typically thought to standardly express purely evaluative concepts. There is a traditional—but controversial—distinction made in normative philosophy between *thin* evaluative concepts, which are purely evaluative, and *thick* evaluative concepts, which are partly evaluative and partly descriptive (Williams 1985). Although aestheticians have not typically made this distinction

using this vocabulary, they have often made similar distinctions between "intrinsically or solely evaluative terms" and "evaluation-added terms" (Sibley 1974), or "verdictive judgments" and "substantive judgments" (Zangwill 1995). While 'beautiful' and 'ugly'—at least on one disambiguation of those terms—are commonly thought to fall into the former (thin/verdictive) category, other adjectives such as 'elegant' and 'graceful' are thought to fall into the latter (thick/substantive) category. One potential difference is that thin evaluative adjectives are plausibly more multidimensional—they can be evaluated with respect to a greater variety of criteria, given their lack of descriptive component—than thick evaluative adjectives. So, do people use thin and thick aesthetic adjectives differently?

Methodological:

- In Study 1 and Study 2, participants always responded to the tasks by choosing one from a list of options. Although we consistently replicated Syrett and colleagues' pattern of results with non-aesthetic gradable adjectives, our experimental paradigm can nevertheless seem comparatively less natural and less behavioral. Could some artifact remain in our experimental paradigm?
- Some of our stimuli sets, including the human face sets, consist of various digital manipulations of one original photograph. Could participant responses be an artifact of our method of stimuli construction? Perhaps some participants refused to make selective or comparative judgments when they deemed two stimuli as too similar or indistinguishable. Or perhaps some participants refused to make selective or comparative judgments when they saw two stimuli as depicting one and the same object. (Compare: the request to identify 'the tall one' may seem odd if one is presented with two photographs of one and the same child before and after a growth spurt.)

We sought to answer all these questions in Study 3. In response to one of the methodological questions, this study adopted a revised experimental paradigm. Each adjective / stimuli set is now presented in the following form (fig. 5).

Participants were instructed to pick out an object by directly clicking on the picture of it, before advancing to the next set. This direct clicking procedure more closely mirrors that of the procedure used in Syrett's original design in giving the participants a sense of immediacy. The new experimental design is more natural than the old one because participants

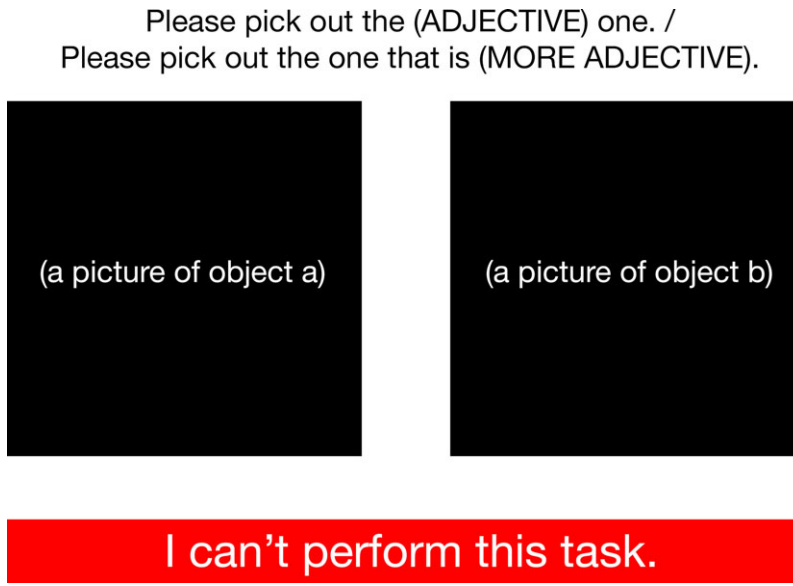

Figure 5. *adjective/stimuli set presentation format for Study 3 & Study 4 (selective phase only)*

can now refuse to perform the task and then give their own reasons.[10] Participants again went through the selective phase and then the comparative phase, and the order of presentation of the adjective / stimuli sets was randomized within each phase.

We also modified the experimental design in response to the other questions. The aesthetic stimuli for this study consisted of photographs of abstract sculptures by Barbara Hepworth, Henry Moore, Constantin Brancusi, and Isamu Noguchi.[11] We chose abstract sculptures to avoid any depiction/depicta confusion; we wanted to make clear to the participants that they have to apply the aesthetic adjectives to the artwork—the sculpture, in this case—and not what the artwork represents. We tested the thin aesthetic

[10] Unfortunately, this revised experimental paradigm also introduces a potential demand characteristic: to fulfill the good participant role, participants may be less inclined to refuse to select an object because that would amount to refusing to complete the purported task. To mitigate this demand characteristic, we used the initial instruction set to hint that refusing to select an object is a perfectly acceptable response. In the instructional set for the selective phase, participants were shown two rods that are bent to different degrees and the adjective 'straight'. Participants were then asked to refuse to pick out the object that is straight and type in the explanation field 'because both rods are bent'.

[11] The specific works presented were: Hepworth, *Oval Sculpture No. 2* (1943/1958); Hepworth, *Image II* (1960); Moore, *Composition* (1932); Moore, *Three Points* (1939); Brancusi, *Endless Column* (1918); Brancusi, *The Bird* (1923/1947); Noguchi, *Red Cube* (1968); Noguchi, *Skyviewing Sculpture* (1969).

adjective 'beautiful' with the Hepworth and Moore sculptures, and the thick aesthetic adjective 'elegant' with the Brancusi and Noguchi sculptures. In response to another one of the methodological questions, all the sculptures look sufficiently distinct from others.

We recruited 40 participants for this study from Amazon Mechanical Turk, again restricting the eligibility to people whose registered location is in the United States and have HIT approval rate greater than or equal to 95%. Prior to any data analysis, we excluded 5 participants for giving at least one obvious-nonsense response.[12] The sample that remains, which was used for subsequent data analyses, consisted of responses from 35 participants (19 women; M_{age} = 31.63; SD_{age} = 9.09).

4.2. Results and Discussion

Again, results from the comparative phase showed a relative prevalence of neither-greater responses with aesthetic adjectives. A decent proportion of participants in Study 3 were unwilling to judge that one object possessed the relevant property to a greater degree with respect to the aesthetic adjectives (fig. 6). We found this tendency with both the thin aesthetic adjective 'beautiful' and with the thick aesthetic adjective 'elegant'. Again, we believe that this phenomenon calls for explanation in its own right.

We then filtered out neither-greater responses from subsequent data analyses. When we compared patterns of compliance and refusal between the adjectives tested, once again we observed the same kind of results that we found in Study 1 and Study 2 (fig. 7). To start, consider the putatively *thin* or *verdictive* aesthetic adjective 'beautiful'. With Hepworth sculptures as stimuli, 'beautiful' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.014, Cramer's V = 0.328) and from 'spotted' ($X^2(1)$ = 33.899, $p < 0.001$, Cramer's V = 0.739). With Moore sculptures as stimuli, 'beautiful' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.002, Cramer's V = 0.406) and from 'spotted' ($X^2(1)$ = 27.279, $p < .001$, Cramer's V = 0.669). Then, consider the putatively *thick* or *substantive* aesthetic adjective 'elegant'. With Brancusi sculptures as stimuli, 'elegant' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.003,

[12] Footnote 8 explains what counts as an obvious-nonsense response. Study 3 adopts a revised different exclusion criterion from Study 2: we now exclude participants rather than individual responses. We have a two-fold rationale. First, we were concerned that excluding just the nonsense responses resulted in response sets that were not the same size for all adjectives tested. Second, we had reservations about the trustworthiness of a participant who gives a nonsense response with at least one adjective / stimuli set. This revision in exclusion criterion was decided prior to conducting Study 3.

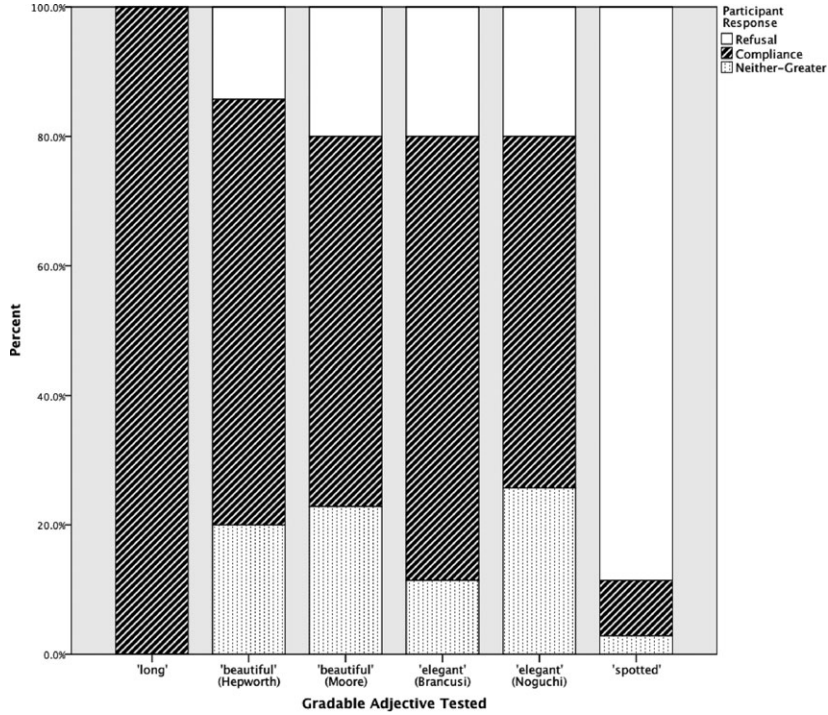

Figure 6. *thin and thick aesthetic adjectives vs. paradigmatic relative and absolute adjectives (refusal/compliance/neither-greater)*

Cramer's $V = 0.366$) and from 'spotted' ($X^2(1) = 31.420$, $p < 0.001$, Cramer's $V = 0.695$). With Noguchi sculptures as stimuli, 'elegant' functioned differently from 'long' (Fisher's exact (2-sided), $p = 0.002$, Cramer's $V = 0.418$) and from 'spotted' ($X^2(1) = 26.139$, $p < 0.001$, Cramer's $V = 0.661$).

In addition to once again demonstrating that aesthetic adjectives do not function like either paradigmatic relative adjectives or paradigmatic absolute adjectives, results from Study 3 bring out other aspects of how aesthetic adjectives functions that hold theoretical interest. First, since the results with 'beautiful' are in line with the results with 'elegant', putatively thin and thick aesthetic adjectives appear to function in more or less the same way. Second, since the results from this study are broadly in line with results from earlier studies, aesthetic adjectives appear to function in more or less the same way with both mundane and artistic objects. We therefore have accumulated further reasons to think that aesthetic adjectives robustly problematize the relative/absolute classification scheme of gradable adjectives.

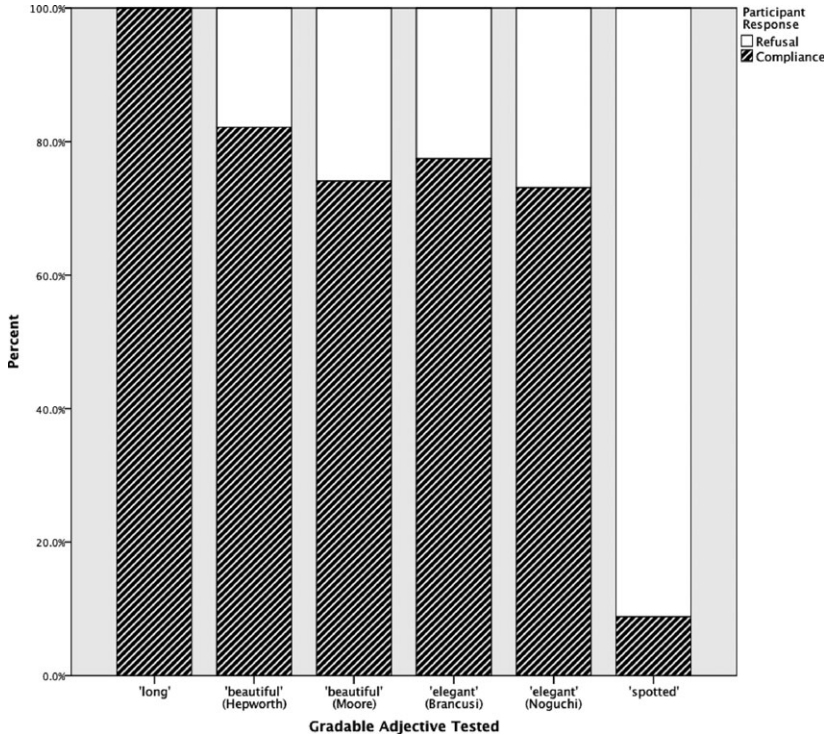

Figure 7. *thin and thick aesthetic adjectives vs. paradigmatic relative and absolute adjectives (refusal/compliance)*

5. Study 4: Two Alternate Hypotheses

5.1. Motivation and Methods

We believe that Studies 1–3 provide strong support for our contention that aesthetic adjectives complicate the relative/absolute classification scheme of gradable adjectives. When we have presented the results of Studies 1–3, a common response is to point to folk relativism about aesthetics as a factor that can help to explain the results. However, it is difficult to pin down a precise hypothesis.

So, in Study 4, we used one simple operationalization to investigate this idea. In the final demographic phase, after all the other tasks are finished, participants were asked *It is commonly said: There is no disputing taste'. Do you agree or disagree?* and responded on a 1 (strongly agree) to 7 (strongly disagree) scale. However, we found no evidence for this factor making any difference in any direction. Participants who refused on the selection tasks with aesthetic adjectives are no more or less likely than participants who complied to be folk relativists about aesthetics, on this measure.

However, in addition, the other aspects of Study 4 investigate more subtle variations on the idea behind the common response. Specifically, two alternative hypotheses have been suggested to us, both of which appeal more indirectly to interpersonal variations in aesthetics.

Hypothesis 1: No Crisp Judgments

According to Kennedy (2011), there is a phenomenon that is distinctive of relative, but not absolute, adjectives: if two objects A and B are judged to be very similar with respect to the relevant property (e.g., very close in size, very close in age), and the relevant adjective is relative, then implicit comparisons such as 'A is the big one' and 'B is the old one' will be judged to be infelicitous even when explicit comparison such as 'A is bigger than B' are fine. So, for example, it is allegedly infelicitous to call one of a pair of children 'the old one' if their birthdays differ by only one or two days. Call this phenomenon *no crisp judgments*.

No crisp judgments could explain the pattern of responses we observed with aesthetic adjectives as follows. It is plausible that there is a significant degree of interpersonal variation in aesthetic attribution. If some subjects judge a pairs of stimuli to be quite similar aesthetically, they may be unwilling to judge either one of the stimuli as 'the beautiful one', 'the ugly one', or 'the elegant one' because of the no crisp judgment phenomenon. But other subjects who judge the pairs to be substantially different aesthetically will have no trouble making such judgments.

Hypothesis 2: Absolute with Varying Thresholds

Although we started this investigation thinking that aesthetic adjectives are likely to be relative, based on the semantic tests that Kennedy presented, philosopher Mark Phelan raised a radical alternative in his comments on our paper at the 2014 Southern Society of Philosophy and Psychology meeting. According to Phelan, aesthetic adjectives such as 'elegant' and 'beautiful' are absolute adjectives with minimum thresholds that vary interpersonally.[13] On Phelan's hypothesis, what happens is that some participants comply with the selection task because only one object meets the minimum threshold of, say, beauty, and other participants refuse because both objects meet the minimum threshold of, say, beauty.

[13] Phelan presented a pilot study, omitted here due to space considerations, that provided some initial support for this hypothesis. We have deposited Phelan's pilot study in the data and material repository associated with our paper, accessible at https://osf.io/6uztd/. We thank him for his permission to share it.

As an analogy, consider how people would respond to the selection task with the minimum standard absolute adjective 'spotted' when presented with discs of one and five spots respectively. Presumably "one spotters" (i.e., those for whom the minimum threshold for counting something as spotted is one) would refuse to pick out 'the spotted object' because they would count both discs as spotted, while "two spotters" would comply because they would count only one of the discs as spotted.

We sought to address these two alternative hypotheses in Study 4. We used the same set of stimuli as the one in Study 3. The selection phase of this study is the same as the one in Study 3 (fig. 5). However, we modified the comparative phase of this study. Instead of asking participants to explicitly compare two objects, we followed one of Phelan's suggestions and asked participants to rate the two objects individually, each on a sliding scale (an example is in fig. 8).

Using a fine-grain measure in the comparative phase allows us to assess both of the hypotheses mentioned earlier. Regarding Hypothesis 1, we can

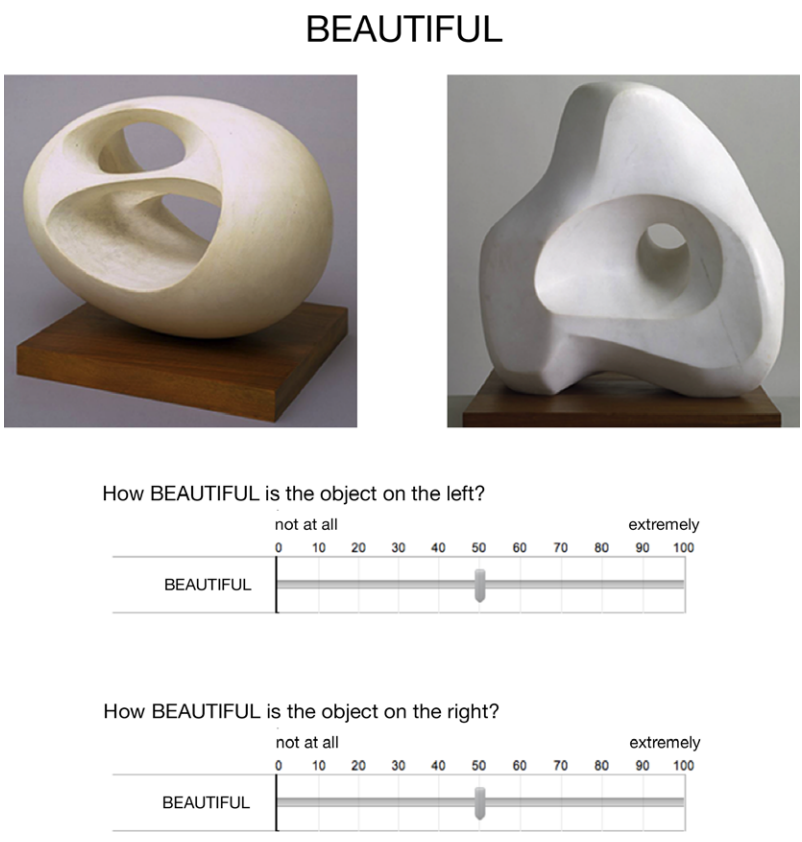

Figure 8. *an example of the adjective / stimuli set presentation format for Study 4 (comparative phase only)*

check whether instances of refusal in the selection phase are accompanied by very small differences in the degree ratings of the respective two objects. Regarding Hypothesis 2, we can check whether some participants are complying with the request to choose the beautiful one simply because only one of the two objects meets the minimum threshold for beauty.

We recruited 50 participants for this study from Amazon Mechanical Turk, restricting the eligibility to people whose registered location is in the United States and have HIT approval rate greater than or equal to 97%. Prior to any data analysis, 8 participants were excluded for failing a comprehension test or for giving one obvious nonsense response. The sample that remains, which was used for subsequent data analyses, consisted of responses from 42 participants (20 women; M_{age} = 39.62; SD_{age} = 13.85).

5.2. Results and Discussion

As before, we filtered out neither-greater responses from subsequent data analyses so that we can again test to see whether the aesthetic adjectives 'beautiful' and 'elegant' are relative or absolute. When we compared patterns of just compliance and refusal between the adjectives tested, once again we observed the same kind of results that we found in Studies 1–3. To start, consider 'beautiful'. With Hepworth sculptures as stimuli, 'beautiful' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.009, Cramer's V = 0.325) and from 'spotted' ($X^2(1) = 38.270$, $p < .001$, Cramer's V = 0.710). With Moore sculptures as stimuli, 'beautiful' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.001, Cramer's V = 0.380) and from 'spotted' ($X^2(1) = 33.740$, $p < 0.001$, Cramer's V = 0.654). Then, consider 'elegant'. With Brancusi sculptures as stimuli, 'elegant' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.049, Cramer's V = 0.237) and from 'spotted' ($X^2(1) = 48.149$, p < 0.001, Cramer's V = 0.776). With Noguchi sculptures as stimuli, 'elegant' functioned differently from 'long' (Fisher's exact (2-sided), p = 0.009, Cramer's V = 0.317) and from 'spotted' ($X^2(1) = 39.336$, $p < 0.001$, Cramer's V = 0.715).[14] Overall, there was a clear replication of the effect observed in Study 3.

But what about the two alternative hypotheses we discussed earlier? In short, we found no clear evidence for either.

To operationalize the no crisp judgments hypothesis, we examined the mean of the difference between the two ratings in the comparative phase, with respect to the participants who refused in the selection phase. For those participants, the mean differences were: M_{diff} = 10.50 for 'beautiful'/Hepworth, M_{diff} = 14.55 for 'beautiful'/Moore, M_{diff} = 12.33 for 'elegant'/

[14] As before, we tested two sets of stimuli with the paradigmatic relative adjective 'long'. We used the one with the smaller difference of length between stimuli in these analyses.

Brancusi, and M_{diff} = 9.63 for 'elegant'/Noguchi. These are not huge differences, but they are not negligible on a 100-point scale either. Although we cannot definitively disprove the no crisp judgments hypothesis, owing to the fact that what counts as a relatively small difference is vague or highly theory dependent, we think the mean differences observed in this study make this hypothesis less plausible.

To assess the absolute with varying thresholds hypothesis, it will be useful to take participants' responses to 'spotted' as a comparison. When participants saw a disk that does not meet the minimum threshold for being spotted—that is, it has 0 spots—participants tended to say that it is not at all spotted (M_{min} = 0.07). However, when we looked at the instances in which participants complied with aesthetic adjectives, the pattern is different. There, the analogous means of the object that participants judged to be less beautiful or elegant indicated that participants still think of them as above their own minimum threshold. The lower means were: M_{min} = 31.27 for 'beautiful'/Hepworth, M_{min} = 21.08 for 'beautiful'/Moore, M_{min} = 27.94 for 'elegant'/Brancusi, and M_{min} = 29.15 for 'elegant'/Noguchi. (Remember the scales go from not at all (beautiful/elegant) to extremely (beautiful/elegant). So, essentially, participants are asked to align their own minimum threshold to 0.[15]) These results suggest that aesthetic adjectives do not function as absolute adjectives with varying thresholds.

6. Classifying Gradable Adjectives, Redux

6.1. Dilemma for the Relative/Absolute Classification Scheme

Let us return to the relative/absolute classification scheme of gradable adjectives, which says that there are two distinct categories of gradable adjectives: *relative* ones such as 'tall' and 'fat', which have context-sensitive standards, and *absolute* ones such as 'straight' and 'full', which do not have context-sensitive standards. As things stand, the relative/absolute classification scheme of gradable adjectives receives support from both Syrett and colleagues' experimental results and somewhat equivocal data from other semantic disagnostics (Kennedy 2007; Toledo and Sassoon 2011; Liao, McNally, and Meskin forthcoming). We will argue that our experimental results raise a dilemma for proponents of this classification scheme.

The first horn of the dilemma: If those proponents accept our experimental methodology, then the strange behaviors of aesthetic adjectives that we

[15] However, as an anonymous referee points out, it is an empirically open question whether participants in fact align their thresholds to 0 on the scale provided. Specifically, the referee notes that the threshold implicitly used in a selective judgment might differ from the threshold used in explicitly responding to a scale, in the comparative phase of Study 4. We think this is a reasonable worry and thank the referee for raising it.

have observed in Studies 1–4 suggest that the relative/absolute classification scheme has to be significantly refined, at the very least. Our studies consistently showed that aesthetic adjectives of various kinds—positive and negative, thin and thick—do not function like either paradigmatic relative adjectives or paradigmatic absolute adjectives. Perhaps the categories of relative and absolute are not exhaustive, and one or more additional categories are needed. Or perhaps the categories of relative and absolute adjectives do not correspond to robust kinds. Or, perhaps, as Liao, McNally, and Meskin (forthcoming) suggest, there may be distinct aspects on which aesthetic adjectives are similar to relative and to absolute adjectives; that is, the relative/absolute classification scheme in fact captures multiple aspects of gradable adjectives' typology, which have not yet been properly delineated.[16]

The second horn of the dilemma: If those proponents of the gradable adjective reject our experimental methodology, then they lose a key piece of evidence in favour of the relative/absolute classification scheme. The significance of our experimental results would be downgraded if one can argue that experimental methods are less suitable than non-experimental semantic diagnostics to uncover semantic properties. However, making this line of argument would also force proponents of the relative/absolute classification scheme to similarly downplay Syrett and colleagues' experimental results. They would then be left with only somewhat equivocal data from other semantic diagnostics.

6.2. Routes for Escaping the Dilemma

We now consider two ways that a proponent of the relative/absolute classification scheme might attempt to make sense of our data. Both, like the two hypotheses considered in Section 5, rest on the idea that there are probably more interpersonal variations in the application of aesthetic adjectives than in the application of descriptive adjectives such as 'tall' and 'spotted'.

An examination of the philosophical tradition suggests that aesthetic adjectives such as 'beautiful' and 'ugly' may exhibit a great deal of polysemy. Kant (1790/1987), for example, distinguishes between "free" or "pure" beauty on the one hand and "dependent" or "adherent" beauty on the other. Sibley (2001c) distinguishes between narrow senses of 'beautiful' and 'ugly' (in which they refer to one among many positive or negative

[16] One anonymous referee suggests that aesthetic adjectives might have especially strong *default* comparison classes that make them less context-sensitive than typical relative adjectives. On this suggestion, aesthetic adjectives are distinct from paradigmatic relative and absolute adjectives, but the distinction is not sharp. Indeed, we explore a variation of this suggestion in Liao, McNally, and Meskin (forthcoming). However, we note that this suggestion by itself would not be sufficient to explain the pattern of mixed usage we found in the set of studies reported in the present paper. We thank the referee for the suggestion.

aesthetic qualities) and wide senses of those terms (in which they are used to refer generically to anything with positive or negative aesthetic value respectively). Levinson (2011) has recently argued that there is an "irreducible variety of visual beauty". One natural way of making sense of some of these proposals is to assume that 'beautiful' and 'ugly' are each associated with multiple related scales (cf. Glanzberg 2007).

On the first route for escaping the dilemma, aesthetic adjectives like 'beautiful' and 'ugly' are polysemous between relative and absolute readings. (Compare Kennedy 2011 on 'old'.)[17] If a significant number of our participants use 'beautiful' and 'ugly' as relative adjectives but others use them as absolute adjectives, then we would expect the intermediate levels of compliance that we found with these aesthetic adjectives. Some would comply because they were treating the terms as relative adjectives, but some would refuse to comply since they were treating the terms as absolute adjectives.

On the second route for escaping the dilemma, the intermediate levels of compliance is due to the fact that some participants simply cannot determine the appropriate scale to apply with a given set of stimuli (while other participants can).[18] One might consider hypothetical cases in which other adjectives plausibly associated with a wide range of scales, such as 'talented' or 'smart', were tested. Confronted with two individuals and asked to pick the talented one, it may be difficult to determine the relevant scale of talent to employ. This version of the polysemy reply, which appeals to the multidimensional nature of evaluative aesthetic adjectives, may also have greater potential to explain our finding of a relative prevalence of neither-greater responses with aesthetic adjectives. Indeterminacy with regard to the relevant scale might have generated participant resistance to judging one object as more ugly than the other.

We believe that these polysemy replies, especially the latter one, are live possibilities. However, they both face the challenge of explaining the results with 'elegant' from Study 3 and Study 4. Although 'beautiful' and 'ugly' are commonly claimed to admit of multiple related scales, we know of no extant discussion of 'elegant' as similarly polysemous, and to the same degree. Given that 'elegant' can plausibly be expected to be less multidimensional than 'beautiful' because it contains a descriptive component, one should expect to find a different pattern of response with 'elegant' than with 'beautiful'. However, the experimental results do not conform to this expectation.

---

[17] Perhaps the earlier discussion of apparently conflicting intuitions about 'pretty' in footnote 4 offers some limited support for this version of the polysemy hypothesis.

[18] We thank Tim Sundell for this suggestion.

7. Further Implications and Avenues for Future Research

Our research represents the convergence of several independent lines of inquiry in linguistics, philosophy of language, philosophical aesthetics, and experimental philosophy. Although we have focused on the linguistics upshots thus far, our research also advances the state of debate elsewhere. In this section we briefly describe the other lines of inquiry and point to what our research adds to each.

7.1. Relation to Predicates of Personal Taste

Linguists and philosophers of language have, in recent years, taken a great deal of interest in a nearby class of adjectives (e.g., 'tasty' and 'fun'), which are often called *predicates of personal taste*. But these researchers have had much less to say about aesthetic adjectives such as 'beautiful' and 'ugly'.[19] For some, the avoidance is by design. Lasersohn (2005) explicitly sets aside 'beautiful' to avoid fundamental issues in aesthetics (645). Similarly, Sundell (2011) avoids aesthetic terms so as to "set aside for the moment as much philosophical baggage as possible" (268). For others, the absence is due to convenience. As Stephenson (2007) acknowledges, "Ultimately it would be desirable to give all of these classes [including taste predicates and aesthetic adjectives] a unified treatment, but (like Lasersohn) I will restrict my attention to the paradigm cases of *tasty* and *fun*" (490). Our experimental paradigm can be extended to test the semantic properties of this class of adjectives. In turn, our results can play a part in either aiding the construction of the elusive unified treatment of aesthetic adjectives and predicates of personal taste, or suggesting a principled distinction between the two.[20]

Moreover, our results can point to similarities between aesthetic adjectives and other segments of natural language. For example, Hansen and Chemla (manuscript) extended the PAT to color terms and found that they also exhibited intermediate patterns of behavior, similar to ones we found with aesthetic adjectives. Their and our results thus collectively suggest a previously-unexplored similarity between aesthetic adjectives and color terms.

7.2. Relation to Experimental Philosophy

In recent years, the limitations of standard philosophical tools have become apparent, with many debates reaching stalemates that cannot be resolved by, say, appeals to intuition and introspection. In other philosophical domains,

---

[19] For some notable exceptions see Egan (2010) and Baker (2012).

[20] McNally and Stojanovic (forthcoming) argue that aesthetic adjectives are distinct from predicates of personal taste on the basis of other semantic diagnostics.

experimental philosophers have done much to generate new debates or move existing ones forward. The same cannot be said for aesthetics.[21] Of the published works of experimental philosophical aesthetics, two—Kamber (2011) and Cova and Pain (2012)—aim to trace the contours of folk aesthetics. In contrast, two others—Meskin *et al.* (2013) and Liao, Strohminger, and Sripada (2014)—aim to uncover the psychological processes that underlie aesthetic judgments and engagements. Our research constitutes a distinctive venture into aesthetic psychology in its aim to uncover the connection between *making* and *communicating* aesthetic judgments. It thus represents another significant step toward understanding the non-perceptual aspects of philosophical aesthetics through experimental methods.

7.3. Relation to Aesthetic Communication

Philosophical aestheticians have a longstanding interest in the nature and use of aesthetic adjectives—consider, for example, the traditional project of defining key aesthetic adjectives such as 'beautiful', 'sublime', and 'ugly'. In the contemporary context, Frank Sibley's influential argument to the effect that the application of aesthetic terms are never solely determined by their non-aesthetic conditions has instigated a significant research programme devoted to exploring those terms and their use in ordinary and critical discourse (Sibley 1959; Kivy 1973). To take another example, Kendall Walton's seminal work on the role played by categories in aesthetic judgments raised the question of whether gradable adjectives such as 'tall' or 'small' might serve as models for understanding aesthetic adjectives (Walton 1970). But despite the interest in aesthetic language in general and aesthetic adjectives in particular, philosophical aestheticians have been notably resistant to engaging with the current theories of the semantics of adjectives that are found in linguistics and philosophy of language. Perhaps this is partly due to the tendency of those fields to avoid addressing distinctively aesthetic adjectives. If so, our results should provide philosophical aestheticians a reason to begin to engage with semantic theories. Naturally, our experimental paradigm can also be extended to test other adjectives that are central in aesthetic discourse, such as 'sublime'.

Moreover, if our results are right, they have the potential to shed some light on aesthetic communication. Many philosophers have suggested that various forms of aesthetic communication (for example, the communication of aesthetic judgments or of the mental states underlying aesthetic characterizations) are impossible or, at least, impossible in the absence of first-person

[21] For an overview of the limited extant literature in experimental philosophical aesthetics, see Cova, Garcia, and Liao (forthcoming).

experience.[22] We think that too much of our ordinary communicative practice about the arts would be incoherent if aesthetic communication were impossible in the absence of the objects of our aesthetic interest. On our view, aesthetic communication in the absence of first-hand experience is not impossible, it is just hard. And, on our view, there is no one reason that it is hard—there are a variety of factors that make such communication more difficult than ordinary communication. Our results, which show a significant amount of interpersonal variation in how agents use aesthetic adjectives, provide one piece of the puzzle in explaining the difficulty with aesthetic communication.[23]

References

Baker, Carl (2012). Indexical Contextualism and the Challenges from Disagreement. *Philosophical Studies* 157 (1): 107–123.

Bierwisch, Manfred (1989). The Semantics of Gradation. In *Dimensional Adjectives*, edited by Manfred Bierwisch and Ewald Lang. Berlin: Springer-Verlag.

Cova, Florian, Amanda Garcia, and Shen-yi Liao (forthcoming). Experimental Philosophy of Aesthetics. *Philosophy Compass*.

—— and Nicolas Pain (2012). Can Folk Aesthetics Ground Aesthetic Realism? *The Monist* 95 (2): 241–263.

Egan, Andy (2010). Disputing About Taste. In *Disagreement*, edited by Ted Warfield and Richard Feldman. New York: Oxford University Press.

---

[22] For example, Richard Wollheim's acquaintance principle, which states that aesthetic judgments "must be based on first-hand experience of their objects and are not, except within very narrow limits, transmissible from one person to another" implies scepticism about communication involving aesthetic judgments (Wollheim 1980: 233). For another example, Philip Pettit holds that "the putatively cognitive state one is in when, perceiving a work of art, one sincerely assents to a given aesthetic characterization, is not a state to which one can have non-perceptual access" (Pettit 1983: 25). If Pettit is right, then the communication of the cognitive states underlying aesthetic characterizations is impossible without the recipient having perceptual access to the work.

[23] Shen-yi Liao's work on this paper was supported by a European Community FP7 Marie Curie International Incoming Fellowship, grant PIIF-GA-2012-328977. Material from this paper was presented at Canadian Society for Aesthetics, European Society of Aesthetics, British Society of Aesthetics, Southern Society for Philosophy and Psychology, London Aesthetics Forum, University of Oxford, Oxford Edgar Wind Society, Deusto University, Bucknell University, Kansas State University, University of Leeds School of Design, University of Leeds Centre for Aesthetics, University of Birmingham, University of Edinburgh, Northern Institute of Philosophy, and University of Colorado. Special thanks to Tim Sundell, Isidora Stojanovic, Mark Phelan, Louise McNally, Rob Hopkins, Nat Hansen, Ron Mallon, colleagues at the University of Leeds, and anonymous referees.

Glanzberg, Michael (2007). Context, Content, and Relativism. *Philosophical Studies* 136 (1): 1–29.

Hansen, Nat and Emmanuel Chemla (manuscript). Color Adjectives, Standards and Thresholds: An Experimental Investigation. Unpublished manuscript, University of Reading.

Iseminger, Gary (2004). *The Aesthetic Function of Art*. Ithaca: Cornell University Press.

Kamber, Richard (2011). Experimental Philosophy of Art. *Journal of Aesthetics and Art Criticism* 69 (2): 197–208.

Kant, Immanuel (1790/1987). *Critique of Judgment*, translated by Werner Pluhar. Indianapolis: Hackett, 1987.

Kennedy, Christopher (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy* 30 (1): 1–45.

—— (2011) Vagueness and Comparison. In *Vagueness and Language Use*, edited by Paul Egre and Nathan Klinedinst. Palgrave Studies in Pragmatics, Language and Cognition. London: Palgrave MacMillan.

—— and Louise McNally (2005). Scale Sructure and the Semantic Typology of Gradable Predicates. *Language* 81: 345–381.

Kivy, Peter (1973). *Speaking of Art*. The Hague: Martinus Nijhoff.

Lasersohn, Peter (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy* 28(6): 643–686.

Levinson, Jerrold (2011). Beauty Is Not One: The Irreducible Variety of Visual Beauty. In *The Aesthetic Mind*, edited by Elisabeth Schellekens and Peter Goldie. New York: Oxford University Press.

Liao, Shen-yi, Louise McNally, and Aaron Meskin (forthcoming). Aesthetic Adjectives Lack Uniform Behavior. *Inquiry*.

——, Nina Strohminger, and Chandra Sekhar Sripada (2014) Empirically Investigating Imaginative Resistance. *British Journal of Aesthetics* 54 (3): 339–355.

McNally, Louise and Isidora Stojanovic (forthcoming). Aesthetic Adjectives. In J. Young (ed.), *Semantics of Aesthetic Judgement*, edited by James Young. Oxford: Oxford University Press.

Meskin, Aaron, Mark Phelan, Margaret Moore, and Matthew Kieran (2013). Mere Exposure to Bad Art. *British Journal of Aesthetics* 53(2): 139–164.

Morzycki, Martin (2012) Adjectival Extremeness: Degree Modification and Contextually Restricted Scales. *Natural Language & Linguistic Theory* 30 (2): 567–609.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko (2009). Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology* 45.

Paolacci, Gabriele and Jesse Chandler (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science* 23(3): 184–188.

Pettit, Philip (1983). The Possibility of Aesthetic Realism. In *Pleasure, Preference and Value*, edited by Eva Schaper. Cambridge: Cambridge University Press. 17–38.

Sibley, Frank (1959). Aesthetic Concepts. *The Philosophical Review* 68(4): 421–450.

—— (1965). Aesthetic and Nonaesthetic. *The Philosophical Review* 74(2): 135–159.

—— (1974). Particularity, Art and Evaluation. *Proceedings of the Aristotelian Society*, *Supplementary Volume* 48: 1–21.

—— (2001a). Adjectives, Predicative and Attributive. In *Approach to Aesthetics: Collected Essays on Philosophical Aesthetics*, edited by John Benson, Betty Redfern, and Jeremy Cox. Oxford: Oxford University Press.

—— (2001b). Aesthetic Judgments: Pebbles, Faces and Fields of Litter. In *Approach to Aesthetics: Collected Essays on Philosophical Aesthetics*, edited by John Benson, Betty Redfern, and Jeremy Cox. Oxford: Oxford University Press.

—— (2001c). Some Notes on Ugliness. In *Approach to Aesthetics: Collected Essays on Philosophical Aesthetics*, edited by John Benson, Betty Redfern, and Jeremy Cox. Oxford: Oxford University Press.

Stephenson, Tamina (2007). Judge Dependence, Epistemic Modals, and Predicates of Personal Taste. *Linguistics and Philosophy* 30(4): 487–525.

Sundell, Timothy (2011). Disagreements About Taste. *Philosophical Studies* 155: 267–288.

Syrett, Kristen (2007). *Learning About the Structure of Scales: Adverbial Modification and the Acquisition of the Semantics of Gradable Adjectives*. Doctoral dissertation. Northwestern University, Department of Linguistics.

——, Evan Bradley, Christopher Kennedy, and Jeffrey Lids (2006). Shifting Standards: Children's Understanding of Gradable Adjectives. In *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North America (GALANA), Honolulu, HI*, edited by Kamil Ud Deen, Jun Nomura, Barbara Schulz, and Bonnie Schwartz. Cambridge, Mass: UConn Occasional Papers in Linguistics 4.

——, Christopher Kennedy, and Jeffrey Lids (2010). Meaning and Context in Children's Understanding of Gradable Adjectives. *Journal of Semantics* 27(1): 1–35.

Toledo, Assaf and Galit W. Sassoon (2011). Absolute vs. Relative Adjectives: Variance Within vs. Between Individuals. *Proceedings of SALT* 21: 135–154.

Walton, Kendall (1970). Categories of Art. *Philosophical Review* 79(3): 334–367.

Williams, Bernard (1985). *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.

Wollheim, Richard (1980). *Art and Its Objects* (2nd ed.). Cambridge: Cambridge University Press.

Zangwill, Nick (1995). The Beautiful, The Dainty, and the Dumpy. *British Journal of Aesthetics* 35: 317–329.