



HAL
open science

What Is Wrong with Aesthetic Empiricism? An Experimental Study

Clément Canonne, Pierre Saint-Germier

► **To cite this version:**

Clément Canonne, Pierre Saint-Germier. What Is Wrong with Aesthetic Empiricism? An Experimental Study. *Review of Philosophy and Psychology*, 2024, 10.1007/s13164-024-00744-x . hal-04631340

HAL Id: hal-04631340

<https://hal.science/hal-04631340v1>

Submitted on 2 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

What is wrong with aesthetic empiricism?

An experimental study

Clément Canonne^{1+*} and Pierre Saint-Germier^{1+*}

1- STMS UMR 9912 (CNRS/IRCAM/Sorbonne Université), Paris, France

⁺ The two authors contributed equally to this work

* Corresponding authors: clement.canonne@ircam.fr; pierre.saint-germier@ircam.fr

Keywords

Aesthetics; Experimental philosophy; Aesthetic properties; Thought experiments; Contextualism

Abstract

According to Aesthetic Empiricism, only the features of artworks that are accessible by sensory perception can be aesthetically relevant. In other words, aesthetic properties supervene on perceptual properties. Although commonly accepted in early analytic aesthetics, Aesthetic Empiricism has been the target of a number of thought experiments popularized by Gombrich, Walton, and Levinson, and purporting to show that perceptually indiscernible artworks may differ aesthetically. In particular, this literature exploits three kinds of differences among perceptually indiscernible artworks that may account for aesthetic differences: relative to categories of art, historical provenance, or means of production. Like in all philosophical thought experiments, the reliability of the elicited intuitions remains an empirical question that we address here with the methods of experimental philosophy. Throughout three studies, we show that, in fact, most people do not believe that non-perceptual properties should modulate our evaluation of an artwork's beauty. However, intuitions were much more divided when considering expressive aesthetic properties (such as intensity), and even clearly reversed when considering artistic properties (such as originality or technical achievement). Overall, our studies show that the central intuitions elicited by the classical indiscernibility arguments strongly depend on the class of manipulated properties (expressive aesthetic properties vs formal aesthetic properties; aesthetic properties vs artistic properties) and are thus more suited to refute artistic empiricism than aesthetic empiricism, narrowly construed.

Conflict of interest statement. The authors have no conflict of interest to declare.

Data availability statement. The complete datasets for the three studies reported here are available at <https://osf.io/pjn7c/>.

Ethical approval and consent. Ethical approval for this study was obtained at INSEAD/ Sorbonne University Center for Behavioural Science, Paris, France. All methods were carried out in accordance with their guidelines and regulations. All participants signed an informed consent.

Introduction

Aesthetic Empiricism (AE) is the view according to which only features of artworks that are accessible by sensory perception can be aesthetically relevant. To borrow Gregory Currie's apt characterization, the Aesthetic Empiricist treats an artwork as "a sensory surface: what is aesthetically valuable in a painting can be detected merely by looking at it. Features that cannot be so detected are not properly aesthetic ones" (Currie 1989: 17). AE is often presented as commonsensical (Davies 2003), or at least "seemingly" commonsensical (Walton 1970: 367). What else could be aesthetically relevant to a painting if not how it looks? Who made it, how, and when, might be relevant for a vast number of issues, e.g., art-historical, economical, moral, political, emotional, etc. But as far as aesthetic evaluation is concerned, the unperceivable may seem off-topic. AE was also adopted by many early analytic aestheticians and philosophers of art, especially those with formalist proclivities (e.g., Bell 1914, Prall 1936, Beardsley 1958), and explicitly endorsed by some artists. Nothing sums up the aesthetic empiricist approach to music better than Duke Ellington's motto: "when it sounds good, it is good" (Ellington 1973). Tides, however, have turned. AE has now become a minority view. The main reason for this fall in disrepute is, arguably, the persuasive effect of a certain sort of thought experiment made popular by Ernst Gombrich (1961) and Kendal Walton (1970), and systematically developed by philosophers such as Arthur Danto (1981), Jerrold Levinson (1980), and Gregory Currie (1989), among others. Those thought experiments typically present a pair of perceptually indiscernible artworks in distinct circumstances, such that the aesthetic properties of the perceptually indiscernible artworks vary from one circumstance to another. For instance, Picasso's *Guernica*, perceived under the category PAINTING, is "violent, dynamic, vital, [and] disturbing", but the same artwork, perceived under the (fictional) category GUERNICA (i.e., the category of three-dimensional versions of Picasso's *Guernica*) becomes "cold, stark, lifeless, or serene and restful, or perhaps bland, dull, boring—but, in any case, not violent, dynamic, and vital" (Walton 1970: 347). If the intuitions about the instantiation of aesthetic properties in such hypothetical circumstances are reliable, then AE is in trouble, for those fictional scenarios seem to put before our eyes manifest examples of aesthetically relevant properties not accessible to sense perception, i.e., the category under which an artwork is correctly perceived, in Walton's case.¹

¹ Following Weinberg (2019), our use of the word "intuition" is not meant to be theoretically loaded. We are concerned with the philosophical practice of appealing to verdicts about hypothetical cases, typically involving

But are those intuitions reliable? Such thought experiments typically involve far-fetched scenarios. Consider for example the precise way in which Walton introduces the category GUERNICA:

Imagine a society which does not have an established medium of painting, but does produce a kind of work of art called guernicas. Guernicas are like versions of Picasso's Guernica done in various bas-relief dimensions. All of them are surfaces with the colors and shapes of Picasso's Guernica, but the surfaces are molded to protrude from the wall like relief maps of different kinds of terrain. (Walton 1970: 347).

The category GUERNICA, clearly made up by Walton for the sake of the argument, has little independent plausibility. How would such a category emerge in this society? Should we assume a historical connection with Picasso's original painting? If not, how come we find the colors and shapes of Picasso's "Guernica" on their surfaces? If so, why do they lack an established medium of painting? In any case, the category GUERNICA appears as a very unusual category of art. It is difficult to know from Walton's description what principles are supposed to guide the aesthetic appreciation of guernicas in this hypothetical society and to what extent we can assume they resemble the principles that guide our appreciation of cubist paintings and sculptures. Machery (2017: 113-116) argues that the unusual character of the thought experiment scenarios is a source of unreliability for three reasons: it obscures what facts hold in addition to the facts explicitly stated; it offers opportunities for misrepresentation; and it may cause usually reliable heuristics to fail. As a result, intuitions elicited by thought experiments involving unusual scenarios may be more prone to variation with irrelevant factors, and be overall less reliable. These characteristics clearly apply to Walton's Guernica case, and others as well. For example, in the counterfactual scenario proposed by Gombrich (1961), Mondrian's actual *Broadway Boogie-Woogie*, which displays a "gay abandon" as compared to the severity of Mondrian's earlier output, is presented as the work of the futurist painter Gino Severini, known for colorful and lively paintings of dance scenes. As a work of Severini, *Broadway Boogie-Woogie* would lose its expression of gay abandon (Gombrich 1961: 297). But how could Severini come up with such a painting? Could it be a spontaneous development of Severini's artistic ideas? Should we rather assume it is a pastiche of Mondrian? Wollheim (1980: 43) and Zangwill (1999: 95) both confess having difficulties imagining this precisely, which opens up space for the issues mentioned by Machery.

either hypothetical works of art, or actual works of art placed in hypothetical contexts, as a kind of data to serve as evidence for and against philosophical theories. "Intuition" is used here as an umbrella term for "whatever that may be" that underlies the case verdicts.

Another issue is that not all intuitions are universally shared, among professional philosophers at least, even though they may be widely shared. Levinson (1980) considers a hypothetical performance of the *Hammerklavier* sonata on a Perfect Timbral Synthesizer, capable of producing sounds timbrally indiscernible from those of a piano, including sounds that are too low or too loud to be generated by a physical piano. He claims that such a performance would lack the sublimity and cragginess that characterizes the sonata when played on a real piano because it forces the performer to struggle with the physical limits of the instrument. Dodd (2007) simply denies that a recording of the *Hammerklavier* on such a Synthesizer necessarily lacks the sublimity and cragginess Levinson ascribes to its piano rendition. According to Dodd, for the sublimity and cragginess of the sonata to be instantiated, it is sufficient that the performance sounds *as if* played on a real piano, and this does not require that it should be played on a real piano (Dodd 2007: 43).² Again, this is not an isolated case. For instance, Zangwill simply rejects Walton's intuition that Picasso's *Guernica* ceases to be violent and dynamic when perceived under the category GUERNICA, even though he admits that it may be *less* violent and *less* dynamic than some three-dimensional guernicas, perceived under the same category (Zangwill 1999: 94).

How stable the empiricist intuitions are in these kinds of cases thus remains an empirical question. This is however precisely the sort of question that the methods of Experimental Philosophy allow to address. The main purpose of this paper, accordingly, is to shed new light on the soundness of such thought experiments by directly investigating the extent to which they elicit stable anti-empiricist intuitions. Given that the demise of AE was in large part the effect of such thought experiments, it is important to make sure that their results are sufficiently robust to justify the rejection of AE. If it turns out that the results of such thought experiments are replicable, then the anti-AE arguments will be reinforced. If not, some new prospects may emerge for the Aesthetic Empiricist and further work may be required from their opponents to conclusively refute the view.

Now, the proliferation of thought experiments meant to refute AE in the literature has slightly complicated the dialectics of the debate. As philosophers created their own variants of Gombrich's and Walton's thought experiments, they came to invoke different kinds of non-perceptual aesthetically relevant properties. Gombrich's Mondrian-Severini case involves

² If Levinson objects that the sublimity and the cragginess are meant to apply to the full performance, including the visible gestures of the performer, then the thought experiments fail as an argument against AE for there will plausibly be perceptible differences between the gestures of a pianist, and the gestures of a performer playing on the Perfect Timbral Synthesizer. If the physical mechanisms of the two instruments are not exactly the same, we should indeed expect physical, and hence visual differences, between the gestures of the performer.

proveniential properties (i.e., properties related to the history of the making of an artwork³), Walton's Guernica case appeals to *categorical* properties (i.e., related to the art category an artwork belongs to), whereas Levinson *Hammerklavier* case showcases *generative* properties (i.e., related to how an artwork is generated). So we have (at least) three kinds of anti-empiricist intuitions. And we cannot assume a priori that, for example, anti-empiricist intuitions pumped by categorical differences stand or fall together with anti-empiricist intuitions pumped by proveniential and generative differences. To have a fair assessment of the potential recalcitrance of empiricist intuitions, we consequently need to take into account this diversity of threads and address empirically the force of each kind of anti-empiricist intuition. This is essentially what we did in the sequence of studies reported below. However, before delving into the details of our experiments, and in order to justify some of our methodological choices, we need to clarify the content and scope of AE, as we shall understand it in this paper.

The nature and scope of AE

Even if the general idea behind AE is relatively clear, its precise formulation requires a number of clarifications. Following Currie, who popularized the label (1989: 17), we shall understand AE as a Supervenience Thesis, according to which aesthetic properties supervene on perceptual properties. To say that a class *A* of properties supervenes on a class *B* of properties is to say that it is impossible for two objects to be indiscernible with respect to properties in the class *B* and yet be discernible with respect to properties in the class *A*. In slogan form: there can't be an *A*-difference without a *B*-difference. For instance, shapes in a digital picture supervene on the colors displayed by the pixels because there can't be a shape difference without a pixel difference. Once the digital artist has assigned a color to each pixel, they have thereby determined the shapes that appear on the picture. And if they want to change one of those shapes, they need to change the color of at least one pixel. Similarly, AE claims that there can't be an aesthetic difference without a perceptible difference. Once the perceptual properties of an artwork are fixed, there is no room for its aesthetic properties to vary, and if the artist wants to change the aesthetic content of the work, they need to change its perceptible content. Since the classic thought experiments mentioned above all involve possible situations in which aesthetic differences are found without any perceptible difference, they all address this supervenience claim.

³ This apt term is borrowed from Davies (2009).

Several things need to be clarified in order to see more precisely how those thought experiments address AE, considered as a supervenience thesis. First, the modal force of supervenience claims, i.e., the meaning of “can’t” in the slogan above, can be understood in many ways, and AE is no exception. For instance, on a weak reading, to say that aesthetic properties supervene on perceptual properties is to say that:

(AE⁻) there is no possible world where two objects are perceptually indiscernible, but aesthetically distinct.

As noted by Currie (1989: 23), if AE amounted only to this, Walton’s *Guernica* case would be irrelevant, because Picasso’s *Guernica* and the hypothetical society of *guernica* makers belong, *ex hypothesi*, to distinct possible worlds. The supervenience claim addressed by Walton’s thought experiment is rather that:

(AE⁺) any two possible objects (possibly belonging to distinct possible worlds) that are perceptually indiscernible are also aesthetically indiscernible.

An important consequence of (AE⁺) is that the contexts surrounding the perceptually indiscernible objects in distinct possible worlds do not matter. In particular, the history of the two objects can vary as much as you like, but if they remain perceptually indiscernible, they should remain aesthetically indiscernible, according to (AE⁺). And this is as it should be: the only thing that matters to Aesthetic Empiricism is the “sensory surface” of the artwork, not its history or any other aspect of the surrounding context. So we should be allowed to compare perceptually indiscernible objects coming from distinct possible worlds when we want to evaluate AE: if they turn out to be aesthetically discernible, then AE fails, if not AE stands. As a result, we shall take (AE⁺) as our official analysis of AE.

In addition to the modal force of the supervenience relation, the scope of AE also depends on the way the classes of supervenient (i.e., aesthetic) and subvenient (i.e., perceptible) properties are understood. We shall address these in turn.

Aesthetic properties are notoriously hard to define. For present purposes, a general characterization will suffice. As a first element of characterization, we may reproduce Sibley’s influential list of expressions typically used to express aesthetic properties: the adjectives “tightly knit”, “deeply moving”, the nouns “balance”, “serenity”, “repose”, and the noun phrase “exciting tension” (Sibley 1959: 421).⁴ The criterion used by Sibley to distinguish the aesthetic use of such terms is that their application requires an exercise of taste from the speaker. Clearly,

⁴ It should be remembered that aesthetic properties are expressed by predicates in a given use (Sibley 1959: 421). Some predicates (e.g., « somber », « vulgar ») may not always be used to express aesthetic properties, even if they typically do in the context of the critical discussions of artworks.

the predicates used by Walton (“violent”, “dynamic”, “vital”, “disturbing”, “cold”, “stark”, “lifeless”, “serene”, “restful”, “bland”, “dull”, “boring”), Gombrich (“gay abandon”), and Levinson (“sublime”, “craggy”, “stormy”, “awesome”) in their thought experiments belong to the same family.

In his discussion of AE as a supervenience thesis, Currie (1989: 19) proposed a slightly different characterization of aesthetic properties as “those properties that can be cited as reasons for an aesthetic judgment (a judgment as to the quality of the work)”. Indeed, expressions in Sibley’s list are often used aesthetically in the context of giving reasons for overall judgments about the quality of artwork. The problem, however, is that the value we might be interested in when we assess the “quality” of artworks sometimes seems to go beyond their sole aesthetic value. Some authors make a distinction between aesthetic and artistic value (Kulka 1981), and as a result, one may acknowledge a distinction between the aesthetic and the artistic properties of artworks, the latter playing a specific role in the justification of judgments about artistic value⁵. According to Levinson, for instance, “artistic properties differ from aesthetic properties in that they do not merely *depend* on the artwork’s relation to other artworks and the surrounding artistic background – they are not merely the perceivable upshot of that contextual placement – but are inherently *a matter of* that relationality” (Levinson 2011: 182-183). Originality and influence are paradigmatic artistic properties in that sense since they essentially depend on relations between the work to which they are applied and other works in the relevant art-historical context. A crucial question is thus whether we should include such artistic properties in the class of supervenient properties relevant to AE.

In many cases, assessing the originality of a work of art requires an exercise of taste. This is so each time a work is original partly *in virtue of* its aesthetic character. So Sibley’s criterion provides a reason to take at least those artistic properties the detection of which requires taste as relevant to AE.⁶ Some authors nevertheless found the inclusion of inherently relational artistic properties in the class of supervenient properties relevant to (AE+) problematic. The concern here is that a difference in originality between two artworks which is only due to the extrinsic relations they bear to other works may not reflect their intrinsic aesthetic character. For example, it is generally acknowledged that Picasso’s cubism owes a lot to Braque’s experimentations in the same direction. Now imagine a possible world where Braque died in infancy and Picasso painted the same cubist canvasses with no comparable external stimulation.

⁵ See in particular Kulka (1981), Levinson (2011) and Stecker (2012).

⁶ Lopes (2011) provides additional reasons to doubt that there is non-aesthetic artistic value, and hence non-aesthetic artistic properties. See however (Huddleston 2012, Hanson 2013) for responses to Lopes’ argument.

In such a world, Picasso's *Le Joueur de Guitarre* (1910) would be *more* original than it is in the actual world, where it owes a lot to Braque's *Broc et violon* (1909). Yet, it is unclear whether this comparatively higher originality of Picasso's *Joueur de Guitarre*, which is only due here to the absence of Braque's *Broc et violon*, reflects a higher *aesthetic* value, as opposed to a higher *art-historical* value. A natural and coherent response to this sort of case from the Aesthetic Empiricist would be to accept a difference in art-historical value while denying any difference in *aesthetic value* between the perceptually indiscernible paintings in the two possible worlds. As a result, thought experiments that establish only inherently relational artistic differences between perceptual indiscernibles would be dialectically weaker than thought experiments establishing differences concerning intrinsic aesthetic properties.

Currie took this limitation seriously and considered a third type of property that is supposed to be more directly relevant to the intrinsic aesthetic character of works, and less directly dependent on extrinsic historical factors, namely the artistic *achievement* embodied in the work itself (1989: 36). For instance, two perceptually indiscernible artworks may embody distinct achievements, if the capabilities of their respective makers are importantly and relevantly different. For instance, the Overture of a *Midsummer Night's Dream* written by Félix Mendelsohn at seventeen is generally thought to have a special marvelous quality in virtue of this exceptionally early maturity.⁷ The same piece hypothetically written the same year by a composer in his seventies may arguably be less marvelous.⁸ Since by assumption, the place of this Overture in the history of music is unchanged, this sort of case does not involve any art-historical difference. Yet, the Aesthetic Empiricist may reply that the difference in the evaluation of the two pieces is rooted in an extrinsic relation to the generative capacities of the composer. It remains coherent for the Aesthetic Empiricist to accept a difference in artistic achievement while denying any aesthetic difference. The music is as “bright, living, moving” and “radiant”, to borrow Grove's words, in the two possible worlds; it is only more *surprisingly* so in the actual world where it is the music of a seventeen-year-old. The appeal to artistic achievement to suggest an aesthetic difference between perceptual indiscernibles is thus still dialectically limited.

⁷ For instance, the British musicologist George Grove described it as “doubtless the greatest marvel of early maturity that the world has ever seen in music – probably in any art” (Grove 1903: 728).

⁸ Currie's own thought experiment compares Picasso's *Demoiselles d'Avignon* to the perceptually indiscernible painting of a hypothetical child in a Martin society where artistic capacities are much higher than those of humans. In particular, Picasso's artistic capacities rival only those of five-year-old children there. In Currie's own presentation, however, the verdict of the thought experiment is not clear, and the route from there to an argument against (AE+) is only indirect. The resulting argument is moreover convincingly criticized by Davies (2003: 32-38).

Note that these dialectical limitations of art-historical properties and artistic achievement properties evaporate if the target claim is *Artistic Empiricism*, i.e., the view that artistic properties supervene on perceptual properties, as opposed to *Aesthetic Empiricism*. The literature is unfortunately not always crystal clear on this issue. We take for example the fact that Davies explicitly takes *artistic* properties as the relevant class of supervenient properties in the definition of *aesthetic empiricism* (2003: 26) as evidence of this sort of confusion⁹. In any case, we shall remain faithful to Sibley's criterion, at least for the sake of this study, and adopt a liberal understanding of aesthetic properties that includes art-historical and artistic achievement properties, to the extent that their detection requires taste. We simply note at this stage that the dialectical strength of a case against AE based only on art-historical or artistic achievement properties would remain weaker, compared to a case involving aesthetic properties reflecting the intrinsic aesthetic character of artworks.

The definition of the class of subvenient properties, i.e., perceptual properties, also raises its own difficulties. Suppose perceptual properties are defined simply as properties perceived by sensory perception, e.g., seen or heard. Now, most aesthetic properties of concrete objects are also perceived, as pointed out by Sibley (1965).¹⁰ Even if taste is required to see the balance characteristic of Mondrian's abstract paintings, it does not cancel the fact that this balance is *seen*, as opposed to inferred from the perception of shapes and colors. As a consequence, AE incurs the risk of being trivialized if the notion of perceptual property is not carefully restricted: if all aesthetic properties count as perceptible properties, then aesthetic differences will be perceptible differences, and we will never have the chance to observe an aesthetic difference without a perceptible difference. Now, Sibley observes that, even though the aesthetic properties of concrete objects are very often perceivable, they constitutively depend *qua* perceivable properties on more basic perceivable properties (Sibley 1965). For instance, the deep balance in a Mondrian painting constitutively depends on the distribution of colors and

⁹ It is also unclear whether Danto's famous gallery of indiscernibles should be taken as a part of an argument against aesthetic empiricism or artistic empiricism, or the aesthetic theory of art. The following passage suggests that Danto uses Aesthetic Empiricism as an assumption in an argument against the aesthetic theory of art: "aesthetics could not explain why one was a work of fine art and the other not, since for all practical purposes they were aesthetically indiscernible: if one was beautiful, the other one had to be beautiful, since they looked just alike" (Danto 2003: 3).

¹⁰ For obvious reasons, AE is significant only when applied to entities endowed with perceptible properties and for which we have plausible antecedent reasons to think that their aesthetic properties are interestingly related to their perceptual properties. As a matter of fact, AE has historically been mostly discussed in the context of the philosophy of fine arts and music. Literature is not directly concerned since the perceptible properties of literary works do not usually make a crucial difference to their aesthetic properties (except perhaps in modernist and experimental works involving creative uses of typography, essentially). Furthermore, while the aesthetics of mathematical theorems is fascinating in itself, we shall leave it aside here, focusing only on works of painting and music.

shapes, which themselves do not depend on more basic perceivable properties. We shall thus take as our subvenient properties the class of perceivable properties *not themselves dependent upon more basic perceptual properties*. This clearly fits the sense in which the artworks presented in anti-empiricist thought experiments are perceptually indiscernible, even if they should turn out to be genuinely aesthetically different. The supervenience of aesthetic properties thus remains a substantive issue.¹¹

Before turning to our studies, we should note that analyzing AE as a supervenience thesis concerning aesthetic and perceptual properties is not totally uncontroversial. An alternative view is that AE is rather about the supervenience of *aesthetic evaluations* on *experiential properties* (Lamarque 2010, Nanay 2015). While the former supervenience thesis connects different classes of *properties*, the latter connects different classes of *mental events* (i.e., aesthetic evaluations vs perceptual experiences). It is in principle possible that two perceptually indiscernible artworks be *experienced* differently, depending on contextual factors. In particular, the popular view according to which perceptual experience is cognitively penetrable suggests that background knowledge about art may significantly affect the way we perceive artworks (Stokes 2014). If the cognitive penetrability of perceptual experience is assumed, then this latter version of AE can easily accommodate the anti-empiricist intuitions elicited by the thought experiments based on perceptual indiscernibles: if the new belief that *Broadway-Boogie Woogie* is by Severini rather than Mondrian already affects the content of our perceptual experience, then the differing aesthetic evaluation is harmless to the second supervenience thesis. To refute it, one would require a difference in aesthetic evaluation between indiscernible perceptual experiences. Nanay (2015: 1) suggests that the version of AE cashed out in terms of mental events is the most interesting one because the property-based one is too easily refuted, unless one accepts implausibly strict formalist views. However, as we saw, even moderate formalists such as Zangwill resist the easy refutation (Zangwill 1999: 92). Furthermore, if one adopts this new construal of AE, and perceptual experience turns out to be cognitively penetrable in the way suggested above, then thought experiments based on perceptual indiscernibles cannot provide any decisive evidence against AE: if the difference in aesthetic

¹¹ This restriction of the supervenience base to low-level perceptual properties is not motivated by, and in any case independent from, the adoption of a conservative view according to which only low-level properties are represented in perceptual experience. From a liberal perspective according to which high-level properties are represented in perceptual experience, it is even more reasonable to limit the supervenience base to low-level perceptual properties. Accepting high-level properties in the supervenience base would make AE much weaker, and even trivial if the liberal accepts that aesthetic properties are represented in perceptual experience (see Stokes 2018 for attempts towards such a nuanced version of such a liberal view). Thanks to an anonymous reviewer for pressing this issue.

evaluation depends on the background knowledge of unperceived facts, and perceptual experience is cognitively penetrable, then this knowledge may penetrate the perceptual experience, for the reasons given by Stokes (2014), and break the requirement of indiscernible perceptual experiences. The latter version of AE would thus not be defeated.¹² So the latter version of AE is the trivial one after all, if we assume the cognitive penetrability of aesthetic experience. However, whether or not aesthetic experience is assumed to be cognitively penetrable, it seems to us that the classic thought experiments do matter for a relevant and substantive version of AE, namely, the supervenience of aesthetic *properties* on perceptual *properties*. So we shall keep that construal as our official definition of AE.¹³

Having clarified our understanding of AE, we may now safely proceed with the presentation of our experiments, which were precisely meant to investigate the prevalence of intuitions associated with AE.

Study 1

Our first study aimed to assess the resistance of empiricist intuitions to appropriate changes in categorical, proveniential, and generative properties of the sort portrayed in the thought experiments of Walton, Gombrich, and Levinson mentioned above. In order to avoid potential recall effects, we used an original scenario schema in which judges had to assess the *beauty* of musical works in a competition, based on the way they sounded plus some limited non-perceptual background information X. In our scenario, after each member of the jury had given their beauty scores, a mistake was spotted by the competition organizers in the non-perceptual background information X for some candidate work: it turns out that this work is not X, but rather Y, which is also a non-perceptual background information. A debate then ensued between the competition organizers about whether they should allow the judges to change their score, with one arguing that being Y rather than X does not matter for the evaluation of beauty (i.e., the empiricist view) while the other argued that being Y rather than X does matter (i.e., the anti-empiricist view), and participants were asked with whom they agreed more.

Musical competitions in which the evaluation is (at least partially) blind to any contextual information are rather common within our musical practices, for example in popular singing

¹² See Nanay (2015: 3) for a similar conclusion.

¹³ Of course, we do not mean to reject a priori the move consisting in saving AE, appropriately reconstrued in mental terms, with reference to the cognitive penetrability of experience (Lamarque 2010), but this is not the view that will be discussed here.

competitions such as *The Voice*, or in orchestra recruitment procedures (in which instrumentalists play behind a curtain so that judges only have access to the sound of the performance, and do not have any information about the author of the performance they are listening to). As such, we believed that musical competitions would provide our participants with a plausible fictional setting for reflecting on the role of non-perceptual information in assessing an artwork's aesthetic value.

Our vignettes were all built from the following schema¹⁴:

Courtney and Darcy organize a music contest. The members of the jury must evaluate the beauty of a piece of music. To minimize bias in the evaluation, the audition is blind: the musicians are placed behind a curtain, so that the judges have access only to the sound of their performance and to a short presentation text about the music played. Immediately after listening to the piece, the judges rate its beauty on a scale of 0 to 10, using an electronic box. After the vote, Courtney and Darcy notice that an error has crept into the presentation text: the piece of music played by the musicians was not X, but Y.

A debate ensues between Courtney and Darcy: should the judges be allowed to change their score?

Courtney thinks they should, because the new information could change the judges' assessment of the beauty of the piece.

Darcy thinks that it should not be done, because there is no reason why this new information could change the judges' assessment of the beauty of the piece.

Who do you agree with more, Courtney or Darcy?

For each vignette, participants had to answer the question using the following scale (note that the ordering of the scale was reversed for half of the participants):

1. *I fully agree with Courtney and not at all with Darcy.*
2. *I agree much more with Courtney than with Darcy.*
3. *I agree slightly more with Courtney than with Darcy.*
4. *I agree slightly more with Darcy than with Courtney.*
5. *I agree much more with Darcy than with Courtney.*
6. *I fully agree with Darcy and not at all with Courtney.*

¹⁴ Note that the vignettes are translated here in English but were presented to our study participants in French.

By filling the variables X and Y with pairs of properties belonging respectively to the class of categorical, proveniential, and generative properties, we generated three full-fledged vignettes, each of which enabled us to test for the intuitive acceptance of the three brands of anti-empiricism identified above. The pairs we selected were the following:

Categorical properties:

X: “entitled *Symphonie Tragique*”

Y: “entitled *Symphonie Parodique*”

Proveniential properties:

X: “composed in the 21st century”

Y: “composed in the 12th century”

Generative properties:

X: “was a collective composition”

Y: “was a collective improvisation”

We chose to build our vignettes around judgments of beauty, which is arguably the quintessential aesthetic property, at least according to the general characterization of aesthetic properties provided above, and general enough to apply to a wide variety of cases. But a few remarks on the choices we made in operationalizing our three types of non-perceptual properties are also in order. Our problem was to find, for each type of non-perceptual property, values of X and Y that would maximize the chance of pumping an anti-empiricist intuition, in order to match the dialectics of the original thought experiments.

To do so, we used first of all properties that had already been brought forward in the philosophical debate on aesthetic empiricism. Among categories of art, Walton includes “media, genre, styles, form, and so forth” (1970: 339), so it is natural to assume the tragic and parodic to count as Waltonian categories of art in the case of music. Unlike fourteenth-century polyphony or American minimalism, the tragic and the parodic are categories that can be applied to musical works based on their perceptible properties alone, independently of their time and place of composition, following the Waltonian definition of art categories which requires them to be “perceptually distinguishable”. They also come with specific sets of standard, contra-standard, and variable properties. It is standard for the tragic in music, and contra-standard for the parodic, to evoke awe; it is standard for the parodic, and contra-standard

for the tragic to evoke laughter; it is variable for both to involve strings. For proveniential properties, we chose to manipulate the date of composition of the candidate musical work, following contextualists' arguments which make extensive use of thought experiments involving perceptually indiscernible works composed by musicians from different historical contexts (Levinson 1980: 11-13, Davies 2001: 74)¹⁵. As to generative properties, the opposition between improvisation and composition is often mentioned as aesthetically relevant, the aesthetics of musical improvisation being widely considered to contrast with the aesthetics of musical composition because of their distinctive modes of generation (Alperson 1984; Brown 1996; Hamilton 2000; Iseminger 2010; Davies 2011).

Second, we tried, within those limits, to maximize the plausibility of our values for X and Y. For the stories to make any sense, it seemed important that our fictional judges could equally believe that both the original and the new background information could apply to the candidate's work. In other words, we needed to find pairs of properties that they could plausibly apply to the same sonic event (i.e., the performance heard by the judges). Hence our choices: post-modernist minimalist vocal works from the 21st century may superficially resemble 12th century music; the same musical structure can be regarded as belonging to a serious or humoresque genre depending on broader inferences that are made based upon the composition's title; and improvisations can turn out to sound like compositions, as well as compositions to sound like improvisations (Canonne 2018).

Finally, we wanted to be able to control for a possible confounding source for anti-AE intuitions, which is simply the tendency we might have to find non-perceptual information more aesthetically relevant when such information is highly unusual (e.g., learning that a performance was not played on a piano but rather on a Perfect Timbral Synthesizer). We thus created two different versions for each of our three vignettes: in the first, the most usual X feature was modified by the least usual Y feature; in the second, the least usual Y feature was modified by the most usual X feature. When contrasting here usual with unusual features of artworks, we simply mean features that are more or less frequently observed in artworks, relative to mainstream current reception practices: it is much less frequent to hear 12th century music (such as Hildegard of Bingen's music) in concert halls than music composed by our contemporaries; similarly; it is a common observation that humor is much rarer in instrumental

¹⁵ The received view about the mechanism at work in Gombrich's original thought experiment is that our aesthetic judgments are sensitive to what we take as the expressive repertoire of the artist (Wollheim 1980: 48-54). We can manipulate that expressive repertoire by changing either the composer's identity, or the time and place of the compositional process, or both. Following the terminology of Davies (2009), we consider all such manipulations as manipulations of the provenance of artworks.

music than expressions of awe, sadness, and fate (Huron 2008); and it is well known that improvisation has declined in classical music since the 19th century, with composition being taken as the default way to produce music, at least in classical music (Gooley 2018). We thus obtained a total of six vignettes, varying both the type of non-perceptual property (categorical, proveniential, or generative) and the usualness of the revision (the revised information is either usual or unusual).

Participants and procedure

340 participants (age = 25.25; women: 250; men: 55; other or did not wish to report: 35) were recruited for this first study through the INSEAD-Sorbonne Université Behavioural Lab. 119 participants were excluded for failing the attention check¹⁶ or the control questions, resulting in a final group of 221 participants (age = 25.21; women: 182¹⁷; men: 35; other or did not wish to report: 4).

Sample size was determined a priori using G*Power 3.1. It recommended 228 participants for an ANOVA test with 6 groups and a medium effect size (0.3). Note that the INSEAD monitored the participants' answers to the attention check and stopped the recruitment when the desired number of participants was reached. Note also that, for each of our studies, we did replicate the analyses using all participants' data (i.e., without excluding any participant) and that the pattern of results remained the same.

Once the study was over, a lottery randomly selected 15 participants who each received 20 euros.

The study was conducted online using Qualtrics. Participants were first randomly assigned to one of two groups – presenting either the version of our three vignettes in which the revised information was usual or the version in which it was unusual. Those three vignettes (one for each of our three types of non-perceptual properties) were then presented in random order to our participants.

¹⁶ The high exclusion rate is explained by the very stringent attention check we used. Participants were explicitly asked *not* to answer the first question (“Do you play a musical instrument?”) and to write the word “merci” in the response box after the question: “If so, which instrument?”. The control questions were simple questions meant to check that participants had carefully read the vignettes. Most of the participants who passed the attention check also correctly answered the control questions.

¹⁷ Only after the recruitment phase for our three studies was over did we realize that our study samples were strongly biased in favor of participants identifying as women, which is certainly not ideal. It may legitimately be worrying that our samples are not balanced enough to allow direct generalization. However, in the absence of clear previous empirical evidence supporting the idea that gender might influence aesthetic judgments, we have no a priori reason to think that our results and interpretations only extend to the female population.

In short, our study followed a mixed factorial design, with *Usualness* (i.e., whether the new background information is more or less usual than the original background information) as a between-subjects factor, and *Non-Perceptual Property* (i.e., whether the story pertains to Categorical, Proveniential, or Generative properties) as a within-subjects factor.

Results

In order to have an overview of whether our participants would rather side with the empiricist or the anti-empiricist position, we first recoded our data: for each participant, we computed their mean rating over the three vignettes, with a mean rating below or equal to 3.5 being taken to indicate that the participant had overall sided with the anti-empiricist position, and a mean rating over of 3.5 being taken to indicate that the participant had overall sided with the empiricist position. We found that, overall, our participants were more inclined towards the anti-empiricist view: 63,80% of them sided with the anti-empiricist position, while 36,20% sided with the empiricist position. An exact binomial test showed that the difference between the majority and minority groups was significant ($p < 0.001$). A one-sample t-test from the midpoint also showed that participants' mean ratings were significantly lower ($M = 3.094$, $SD = 1.139$) than the threshold value of 3.5 ($t = -5.304$, $df = 220$, $p < 0.001$).

To assess more precisely the impact of our two experimental factors (*Usualness*; *Non-Perceptual Property*) on participants' ratings (from 1 to 6, with the higher meaning closer to the empiricist position), the data were then analyzed through a 2-way ANOVA, using the EZ package in R. Our statistical analysis revealed an absence of main effect of *Usualness* ($df = 219$, $F = 0.682$, $p = 0.410$, $\eta^2 = 0.002$) and a main effect of *Non-Perceptual Property* ($df = 438$, $F = 30.183$, $p < 0.000$, $\eta^2 = 0.057$). No significant interaction between our two factors was found ($df = 438$, $F = 0.337$, $p = 0.714$, $\eta^2 = 0.001$). A post-hoc Tukey test showed that ratings for the "Proveniential" vignettes were 0.376 higher than ratings for the "Categorical" vignettes ($p = 0.026$), and 0.905 higher than ratings for the "Generative" vignettes ($p < 0.001$), and that ratings for the "Categorical" vignettes were 0.530 higher than ratings for the "Generative" vignettes ($p = 0.001$) (see Figure 1a; see also Figure 1b for rating distributions in each one of our conditions). In other words, participants were more likely to allow judges to revise their evaluation of the beauty of the musical work under consideration when the background information targeted the work's generative properties (i.e., whether the piece was improvised or composed), as compared to both categorical (i.e., whether the piece's title was *Symphonie Tragique* or *Symphonie Parodique*) and proveniential properties (i.e., whether the piece was composed in the 12th or in the 21st century). While the emphasis, in philosophical aesthetics and especially in the

discussion of aesthetic contextualism (Levinson 2017), is often put on proveniential properties broadly construed, it would seem that they do not play a particularly significant role in the aesthetic evaluation of artworks, after all. A one sample t-test from the midpoint indeed showed that participants' ratings in the case of our "Proveniential" vignette were not significantly different than the threshold value of 3.5 ($M=3.520$, $SD=1.625$, $t=0.186$, $df=220$, $p=0.852$), suggesting that intuitions were particularly fuzzy in that case. On the contrary, the arguably less-discussed generative properties of artworks seem much more crucial to the way we think about their value ($M=2.615$, $SD=1.428$, $t=-9.212$, $df=220$, $p<0.001$). This result might be related to the effort heuristic – our tendency to rely on the amount of effort used to create something as a proxy for its value (Kruger et al. 2004). It is indeed arguably much more difficult to improvise a given piece than to compose the very same piece (Kania 2008). This may explain why our participants were particularly inclined to let our fictional judges change their rating of the musical piece's beauty when they learned that it was improvised rather than composed (or composed rather than improvised).

Strikingly, the degree of usualness of the new information provided to the judges had no effect on our participants' intuitions: they were equally inclined towards anti-empiricism, whether the new information about the evaluated artwork made it stand out from the ordinary (e.g., being an improvisation) or not (e.g., being a composition). The anti-empiricist intuition thus seems particularly robust, as it holds despite such an important difference in the framing of our vignettes, contrary to what is sometimes observed in experimental philosophy (Weinberg 2019).¹⁸

¹⁸ It may be fair to point out however that our most unusual features are still much more usual than what we find in the original philosophical thought experiments (*guernicas*, Mondrian-likes painted by Severini, perfect timbral synthesizers).

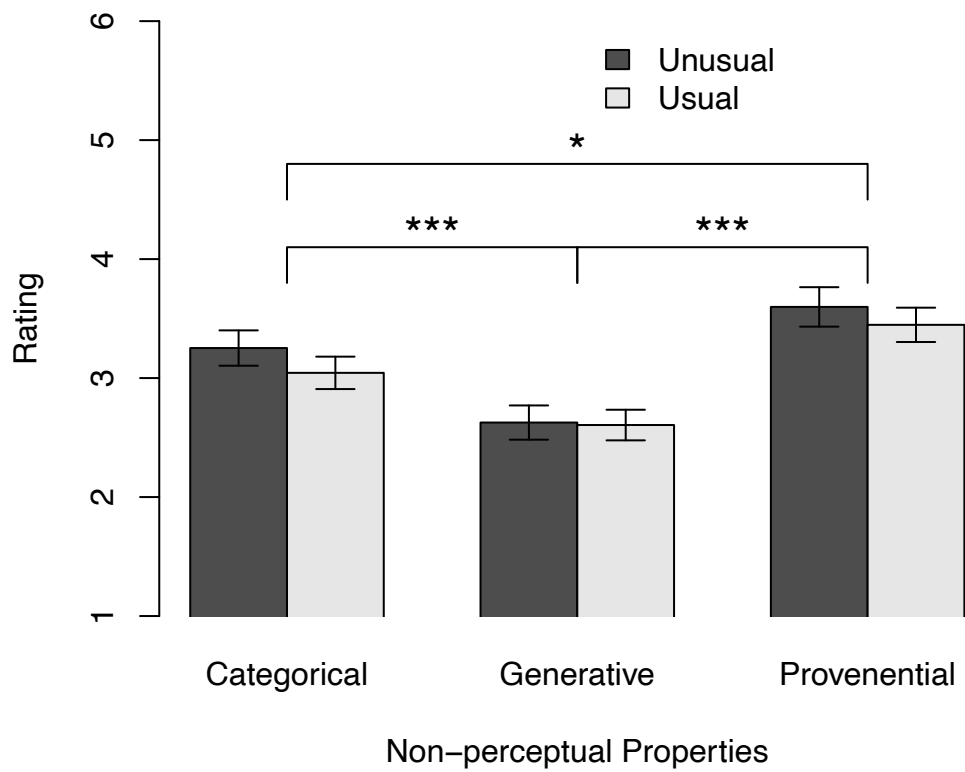


Figure 1a. Results from Study 1. Error bars show standard error (95% interval). * represents the significant outputs of post-hoc tests, with a threshold of $p < .05$; **: $p < .01$; ***: $p < .001$.

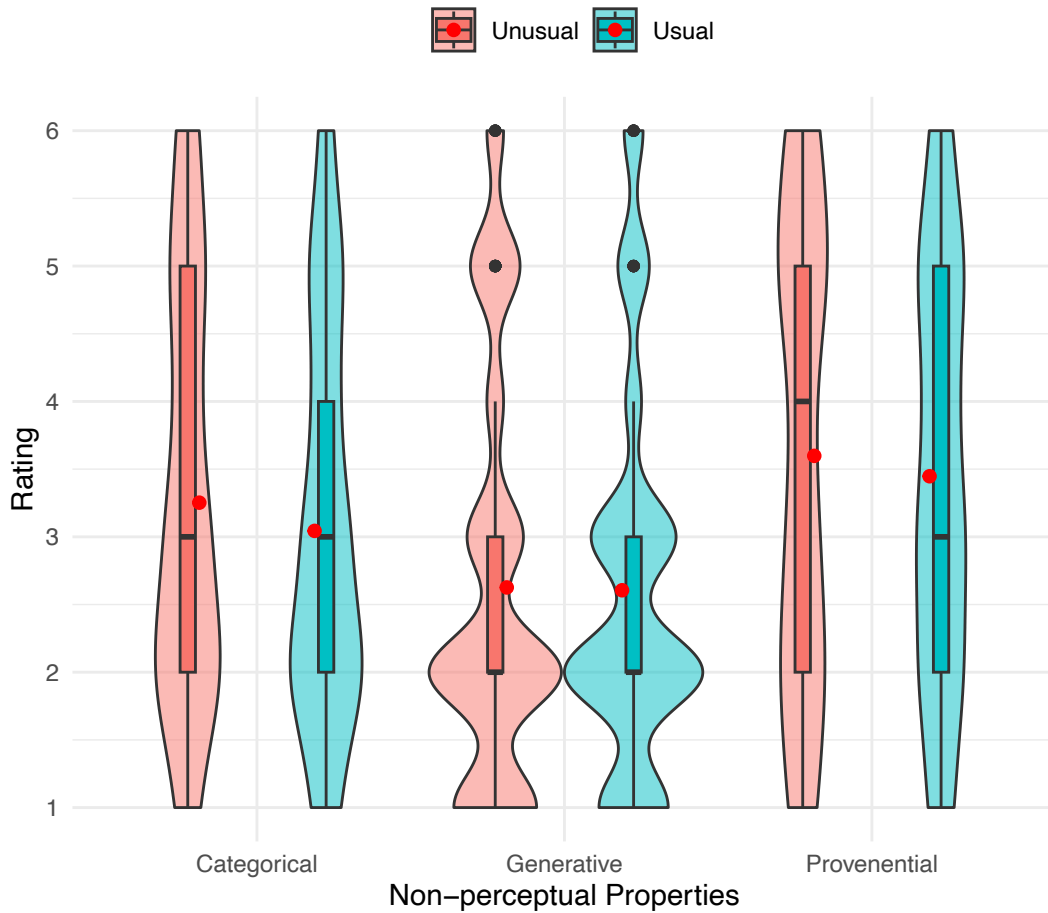


Figure 1b. Distribution of participants' ratings for Study 1. For each condition, the overall shape shows the distribution shape of the ratings (with wider sections representing a higher probability that participants gave that rating), the black horizontal line shows the median, the vertical rectangle shows the interquartile range, the vertical line shows the rest of the distribution (with outliers as black dots) and the red dot shows the mean.

Study limitations

Our first study thus seems to show that anti-empiricist intuitions initially identified by Walton, Gombrich, and Levinson are rather robust, and can be evidenced in structurally similar but different hypothetical cases. However, this study also faces a number of important limitations, which potentially undermine such a straightforward interpretation. First, the anti-empiricist tendency found in our study might be explained by a bias in favor of the correction of mistaken information, independently of the content of that mistaken information. Second, the scenario implicitly presented the unspecified non-perceptual background information given to the judges as relevant to their aesthetic judgment. This may be seen as inducing a bias against empiricism from the start. Third, the musical case might have introduced an unnecessary confusion between the evaluation of the musical performance – to which there is direct sensory access – and the evaluation of the musical work underlying the performance – to which it is debatable whether

there can be direct sensory access or not. Fourth, and most importantly, it did not control for the interpretation of “beauty” between an unspecific reading roughly synonymous with “valuable” when applied to artworks, and a specific reading where beauty is contrasted as a specifically aesthetic property with other artistic qualities we often appreciate in art, e.g., innovation, craftsmanship, but may not, or not only, consist in purely aesthetic properties. It might well be that the setting of a music competition, where artistic merits, rather than purely aesthetic ones, usually occupy the center of the judges’ verdicts, pushed our participants to favor the broader interpretation of the term “beauty” when reading our vignettes, and to think that our fictional judges had to assess the overall artistic value of the musical pieces they heard (which includes but is not limited to its intrinsic aesthetic properties). If that is indeed the case, the observed drive towards anti-empiricism may not be a drive towards *aesthetic* anti-empiricism *per se* but rather towards *artistic* anti-empiricism, which is a distinct issue, as we pointed out already. We thus designed a second study, addressing those four limitations, to see if our results would still hold.

Study 2

For our second study, we adapted the scenario from Study 1 so that no non-perceptual background information was given in advance to the jury. As a result, the relevance of non-perceptual background information was not presupposed and no correction was necessary. The debate between the competition organizers was now to decide whether some surprising property X, contrasting with the more usual property Y should be given to the jury at all. We also switched the setting to a painting competition, in order to avoid potential issues due to the abstract ontological nature of musical works. Finally, our fictional judges were divided into three groups, each in charge of evaluating a specific aspect of the candidates’ work: its beauty, originality, or technical achievement. In framing our scenario in such a way, beauty was now explicitly contrasted with two central varieties of artistic values found in artworks that may plausibly be independently evaluated in the context of a competition, favoring a stricter reading of the term “beauty”. Additionally, contrasting explicitly beauty with originality and technical achievement allowed us to directly contrast aesthetic with artistic evaluation, and thus to investigate whether the elicited intuitions by the kind of thought experiments suggested in the literature are more apt to address Aesthetic or Artistic Empiricism.

All the vignettes were built from the following schema:

Courtney and Darcy organize a painting contest. Artists are invited to send in their paintings and a jury of experts is convened to evaluate them. Courtney and Darcy planned to award three prizes: the prize for beauty; the prize for originality; and the prize for technical achievement. The experts are then divided into three separate committees, each of which is responsible for specifically evaluating the beauty, originality, or technical achievement of the paintings submitted by the candidates, on a scale of 0 to 10. In order to minimize bias in the evaluation, Courtney and Darcy chose not to reveal the identity of the artists and to provide the jury with the minimum information necessary for the evaluation.

All of the proposed paintings appear to be abstract art, which is fashionable among young artists who want to break into the art world. Upon reading the introductory texts provided by the artists, Courtney and Darcy discovered that one of the paintings was in fact X, although neither the title nor the appearance of the painting gave any indication of this, while all of the other paintings were Y. A debate ensued between Courtney and Darcy as to whether or not this information should be given to the various evaluation committees: should each painting be indicated as X or Y?

Three questions were then asked in random order (one for each type of Judged Property Z, i.e., “beauty”, “originality”, “technical achievement”), all of the following form:

Regarding the committee in charge of the Z award, Courtney thinks it is necessary to give this information because being X or Y can change something to the Z of a painting.

Darcy thinks it is not necessary to give this information because being X or Y cannot change anything to the Z of a painting.

Who do you agree with more, Courtney or Darcy?¹⁹

The three pairs of values for X and Y (one for each type of non-perceptual properties) were the following:

Categorical properties:

X = “belongs to the genre of abstract landscape”

Y = “belongs to the genre of monochrome”.

¹⁹ The responses of our participants were then collected by means of the same 6-point Likert scale as above.

Provenential properties:

X = “is painted in 1940 by an artist who is now over a hundred years old”

Y = “is painted in the 2020s by an artist born after 1990”

Generative Properties:

X = “is finger-painted”

Y = “is brush-painted”

Again, some justifications are in order regarding the choices we made in operationalizing our three types of non-perceptual property. For the provenential properties, we selected once again the artwork’s creation date, for the same reasons as mentioned above. For the categorical properties, we took our inspiration from Danto’s famous gallery of indiscernibles (Danto 1981), which contrasts a monochrome with an abstract landscape. Finally, for the generative properties, given that the term “improvisation” is rarely used with reference to paintings, we selected a contrast that more specifically targeted the very material means used in painting the artwork (i.e., a brush *versus* the artist’s own fingers).

Participants and procedure

254 participants (age = 25.58; women: 185; men: 38; other or did not wish to report: 31) were recruited for this second study through the INSEAD-Sorbonne Université Behavioural Lab. Participants recruited for the previous study could not enroll in this second study. 82 participants were excluded for failing the attention check or the control questions, resulting in a final group of 172 participants (age = 25.40; women: 130; men: 33; other or did not wish to report: 9).

Sample size was determined a priori using G*Power 3.1. It recommended 153 participants for an ANOVA test with 9 groups and a large effect size (0.4). Once the study was over, a lottery randomly selected 15 participants who each received 20 euros.

The study was conducted online using Qualtrics. Participants were first randomly assigned to one of three groups – corresponding to our three vignettes (one for each type of non-perceptual properties). Participants had then to answer three questions (one for each type of judged properties), presented in random order to the participants.

In short, our study followed a mixed factorial design, with *Non-Perceptual Property* (i.e., whether the story pertains to Categorical, Provenential, or Generative properties) as a between-

subjects factor, and *Judged Property* (i.e., whether the question asked to the participants is about the evaluation of Beauty, Originality, or Technical achievement) as a within-subjects factor.

Results

First, using the same recoding procedure as in our previous study, we found that a majority of our participants sided with the anti-empiricist position when it came to the evaluation of the artwork's originality (66.86% vs 33.14% for the empiricist position) or of its technical achievement (59.88% vs 40.12% for the empiricist position), which is highly similar to the pattern of results we found in our first study. An exact binomial test showed that the difference between the majority and minority groups was significant ($p < 0.001$ for originality; $p = 0.012$ for technical achievement). A one sample t-test from the midpoint also showed that participants' ratings were significantly lower for both Originality ($M = 2.971$, $SD = 1.550$) and Technical achievement ($M = 3.186$, $SD = 1.651$) than the threshold value of 3.5 ($t = -4.476$, $df = 171$, $p < 0.001$, for Originality; $t = -2.494$, $df = 171$, $p = 0.014$, for Technical achievement). This suggests that anti-empiricist intuitions are indeed largely prevalent when it comes to relational properties such as originality or technical achievement.

However, the results were completely reversed when it came to the evaluation of the artwork's beauty: here, a vast majority of our participants sided with the empiricist position (80.81% vs 19.19% for the anti-empiricist position). An exact binomial test showed that the difference between the majority and minority groups was significant ($p < 0.001$). A one sample t-test from the midpoint also showed that participants' ratings were significantly higher ($M = 4.744$, $SD = 1.322$) than the threshold value of 3.5 ($t = 12.345$, $df = 171$, $p < 0.001$). Such a striking difference between our two studies could suggest that, in our first study, participants were indeed considering "beauty" as a predicate mostly relevant for artistic, rather than aesthetic evaluation, in the context of the competition.

Second, in order to assess more precisely the impact of our two experimental factors (Judged Property; Non-Perceptual Property) on participants' ratings (from 1 to 6, with the higher meaning closer to the empiricist position), the data were analyzed through a 2-way ANOVA, using the EZ package in R. Our statistical analysis revealed a main effect of Judged Property ($df = 338$, $F = 104.826$, $p < 0.000$, $\eta^2 = 0.224$), a main effect of Non-Perceptual Property ($df = 169$, $F = 5.591$, $p = 0.004$, $\eta^2 = 0.034$), and a significant interaction between our two factors ($df = 338$, $F = 3.342$, $p = 0.001$, $\eta^2 = 0.018$). A post-hoc Tukey test showed that ratings for the "Beauty" question were on average 1.773 higher than ratings for the "Originality" question ($p < 0.001$) and 1.558 higher than ratings for the "Technical achievement" question ($p < 0.001$); and that

ratings for the “Categorical property” vignette were on average 0.669 higher than ratings for the “Generative property” vignette ($p < 0.001$) and 0.526 higher than ratings for the “Proveniential property” vignette ($p = 0.004$) (see Figure 2a; see also Figure 2b for rating distributions in each one of our conditions). When looking more precisely at the interactions between our two factors, post-hoc t-tests (using Bonferroni correction to account for multiple comparisons) showed that participants’ ratings for the “Originality” question were significantly higher in the “Categorical property” vignette than in the “Generative property” and “Proveniential property” vignettes (resp. $t = 4.259$, $df = 99.157$, $p < 0.001$; $t = 2.916$, $df = 105.99$, $p = 0.016$). And a similar pattern was found for the “Technical achievement” question, with ratings being significantly higher in the “Categorical property” vignette than in the “Generative property” ($t = 2.727$, $df = 106.23$, $p = 0.028$). Categorical properties thus seem to be much less impactful for the evaluation of an artwork’s artistic value than generative and proveniential properties, which are arguably more directly tied to crucial artistic values such as originality (the historical context in which the work was produced should play a great role in assessing its originality) and technical achievement (the way the work was produced should play a great role in assessing how much of a technical achievement it is). But crucially, there was no such significant difference in the way our three types of vignettes affected participants’ ratings for the “Beauty” question, which were consistently high across the three vignettes. The fact that participants were equally indifferent to generative, proveniential, and categorical properties when it came to judgments of beauty suggests that Aesthetic Empiricism might be particularly robust amongst our subjects, at least for beauty.

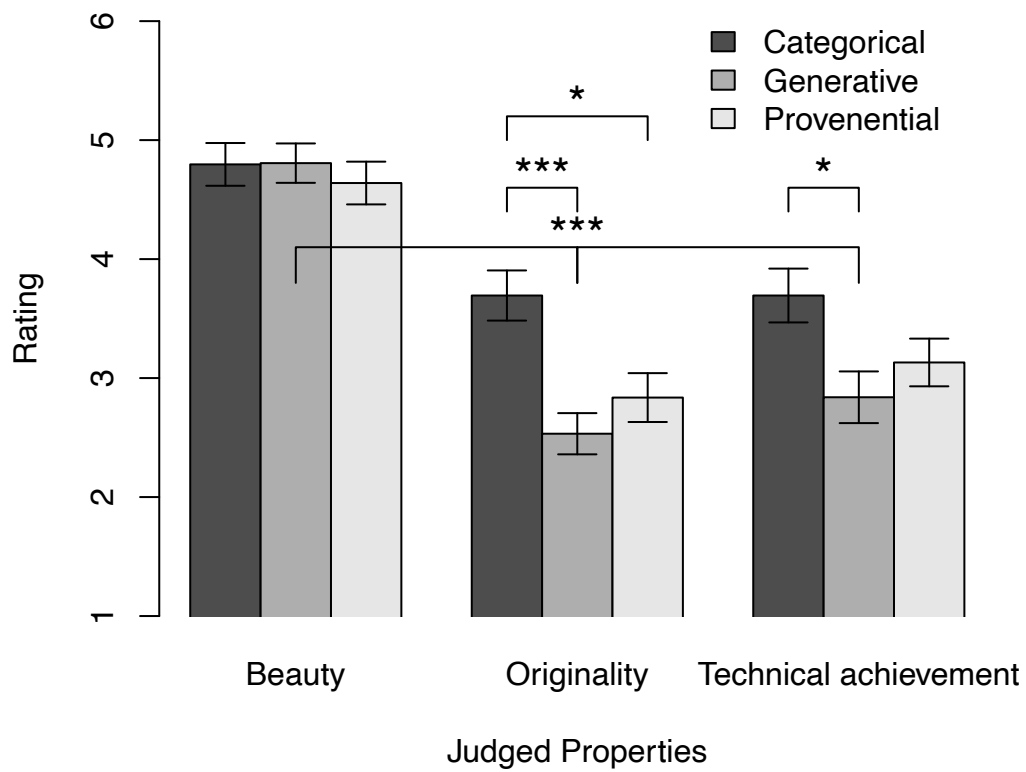


Figure 2a. Results from Study 2. Error bars show standard error (95% interval). * represents the significant outputs of the post-hoc tests, with a threshold of $p < .05$; **: $p < .01$; ***: $p < .001$.

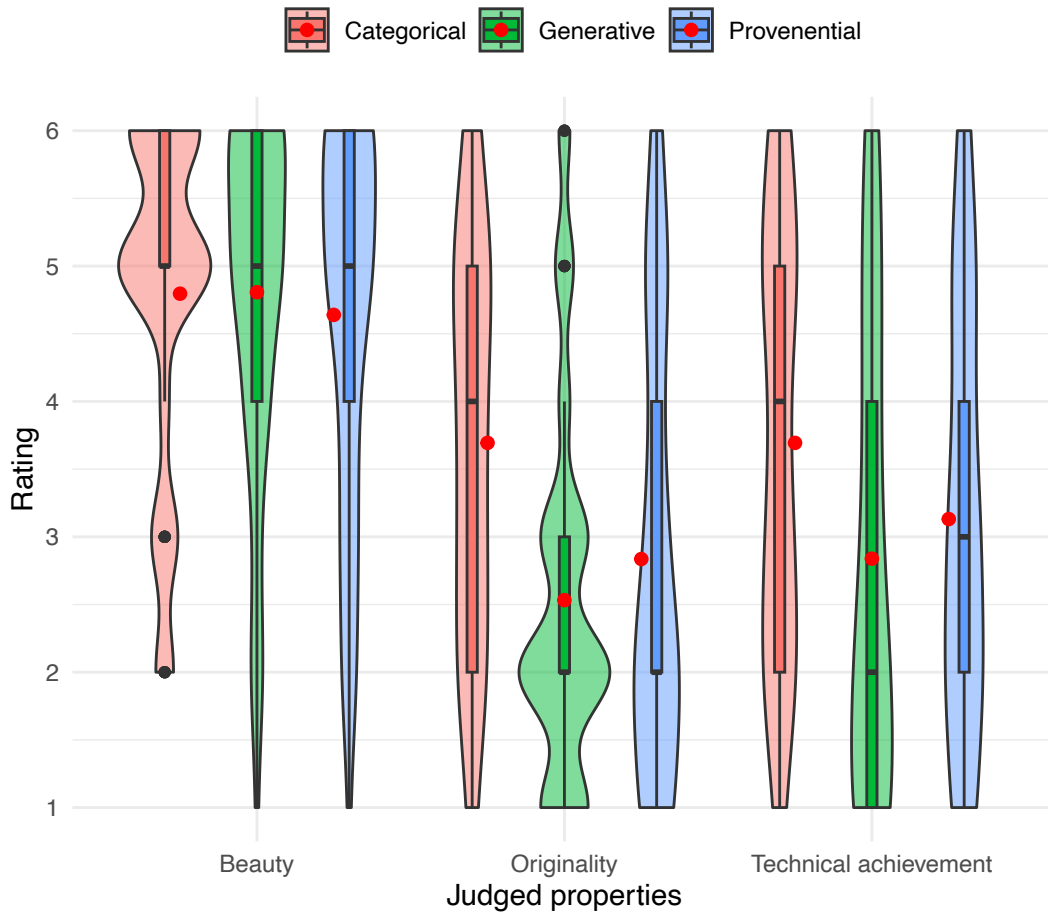


Figure 2b. Distribution of participants' ratings for Study 2.

Study limitations

This second study suggests a totally different picture than our first study: when forcing our participants to endorse an aesthetic reading of the term “beauty”, they were largely inclined to believe that information about a painting’s generative, proveniential, or categorical properties was not relevant in assessing its beauty. In that sense, our participants’ intuitions seem clearly empiricist when it comes to intrinsic aesthetic properties such as beauty – the beauty of a painting being mainly a matter of how this painting *looks* – even though anti-empiricist intuitions were still more prevalent among our participants for relational artistic properties such as originality or technical achievement.

Now, anti-empiricist philosophers could still object that these results do not necessarily undermine their position, for three reasons.

First, given the apparent tension between the results from our first two studies, they might require additional empirical confirmation before taking for granted the prevalence of the empiricist intuition. In particular, they might argue that we have introduced a bias in favor of

AE, by explicitly stating at the outset of our vignettes that providing the judges with the identity of the artists might bias the evaluations and that the diverging result patterns could just be explained by this bias.

Second, they might object that beauty is a much too coarse-grained aesthetic property: in that sense, our results could just be due to the difficulty our participants had in relating the very specific properties described in our vignettes (e.g., whether a painting was painted with a brush or with fingers) with an aesthetic quality as vague and general as beauty. Most thought experiments found in the literature on aesthetic empiricism indeed rely on aesthetic properties that are much more specific than beauty: “vitality” in Walton’s *guernicas*, for example, “gay abandon” in Gombrich’s, or “storminess” in Levinson’s perfect synthesizer case. In particular, the aesthetic properties brought forward in this context seem to have much more to do with *expressive* aesthetic properties than with generic aesthetic properties such as beauty.

Third, the thought experiments found in the literature are generally not as abstract as the cases presented in our own vignettes, as they usually rely on well-known artworks (e.g., Picasso’s *Guernica* or Beethoven’s *Hammerklavier*), which are then contrasted with one of their counterfactual versions. Readers of those thought experiments are thus supposed to have a clear enough representation of how the artwork under discussion actually looks or sounds like, often supplemented by appropriate reproductions allowing them to exercise their own aesthetic judgment. By contrast, in our vignettes, participants lacked any precise idea of how the artworks looked. This is in striking contrast to our usual epistemic situation when we engage with artworks. As a result, the intuitions elicited in our experiments might be equally removed from the judgments made in our actual aesthetic life (see Puy 2021 for similar concerns on experimental approaches to musical ontology; see also Weinberg 2019 on the phenomenon of “filling in” in thought experiments that rely on abstract descriptions rather than concrete examples).

For all those reasons, we decided to conduct one last study that would stick more closely to the thought experiments used by anti-empiricists, presenting participants with representations and sound excerpts of the artworks under discussion, and contrasting the effects of non-perceptual properties on the evaluation of both generic aesthetic properties and expressive aesthetic properties.

Study 3

For our third study, we adapted the thought experiments of Walton, Gombrich, and Levinson, respectively covering categorical, proveniential, and generative properties, to the fictional setting used in our previous studies (i.e., the evaluation of an artwork about which new information arises). This gave us three important advantages: first, we could directly test how relevant those seminal thought experiments seemed to naive subjects; second, we could provide our participants with concrete examples of how the artworks discussed in the vignettes sounded or looked; and third, we could assess the prevalence of anti-empiricist intuitions for three distinct artistic domains (street art, painting, and classical music).

As mentioned above, a crucial characteristic of those thought experiments is that they target expressive aesthetic properties: “vividness” for Walton, “storminess” for Levinson, and “gay abandon” for Gombrich. Abstracting one step up, those three properties can be related to an artwork’s *intensity*. We thus created two versions of each one of our three scenarios: one in which the judges were meant to assess the artwork’s beauty (contrasted with its originality and technical quality); and one in which the judges were meant to assess the artwork’s intensity (contrasted with its originality and technical achievement).

For each thought experiment, we thus constructed a scenario schema where the variable X could be filled with two types of aesthetic property: “beauty” or “intensity”.

Categorical Properties: Guernica

Courtney and Darcy work for a street art festival. In order to make the festival better known, they organize a participative competition. Randomly selected citizens are invited to come to the festival and evaluate the exhibited works according to one of the following three criteria: X, originality, and technical achievement. Concretely, for each work and for each of these criteria, the citizens give a mark from 0 to 10.

Among the works exhibited is a mural entitled Gernikara, reproducing the famous painting Guernica by Picasso on the wall of a disused factory, of which this [Figure 3a] is a photograph.



Figure 3a. Mural reproduction of Pablo Picasso's painting *Guernica* (on tiles, at Gernika-Lumo, 11 Allendesalazar). Photograph by Papamanila. The picture was presented to the participants with no caption.

After all the citizens have completed their evaluation, Courtney and Darcy learn from talking to the artist that Gernikara actually belongs to the (little known there, but very fashionable in Bilbao) genre of guernicas: guernicas are three-dimensional versions of Picasso's painting, whose principle is to use the dimension of depth. Here [Figure 3b] is a recent example of a guernica:



Figure 3b. *Guerniskate*, mural painting by Nils Inne, 2019. © Nils Inne. The picture was presented to the participants with no caption.

Courtney and Darcy quickly agree about whether this new information matters for the evaluation of the originality and technical quality of Gernikara. But they start a dispute about the importance of this new information for the evaluation of the X of Gernikara: can we, or can we not, keep the marks given by the citizens in charge of evaluating the X of Gernikara?

Courtney thinks that we can't keep the marks given by the citizens in charge of evaluating X because considering Gernikara as a fresco or a guernica can make a difference to its X.

Darcy thinks that we can keep the marks given by the citizens in charge of evaluating X because considering Gernikara as a fresco or a guernica cannot make a difference to its X.

Provenential Properties: Broadway Boogie-Woogie

Courtney and Darcy work for a museum. In order to make the museum's collection better known, Courtney and Darcy organize a participatory competition. Randomly selected citizens are invited to visit the museum and evaluate its works according to one of the following three

criteria: X, originality, and technical achievement. Specifically, for each work and for each of these criteria, the citizens assign a score from 0 to 10.

Among the works in the museum's collection is the following painting [Figure 3c], known as Broadway Boogie-Woogie and attributed to Piet Mondrian:

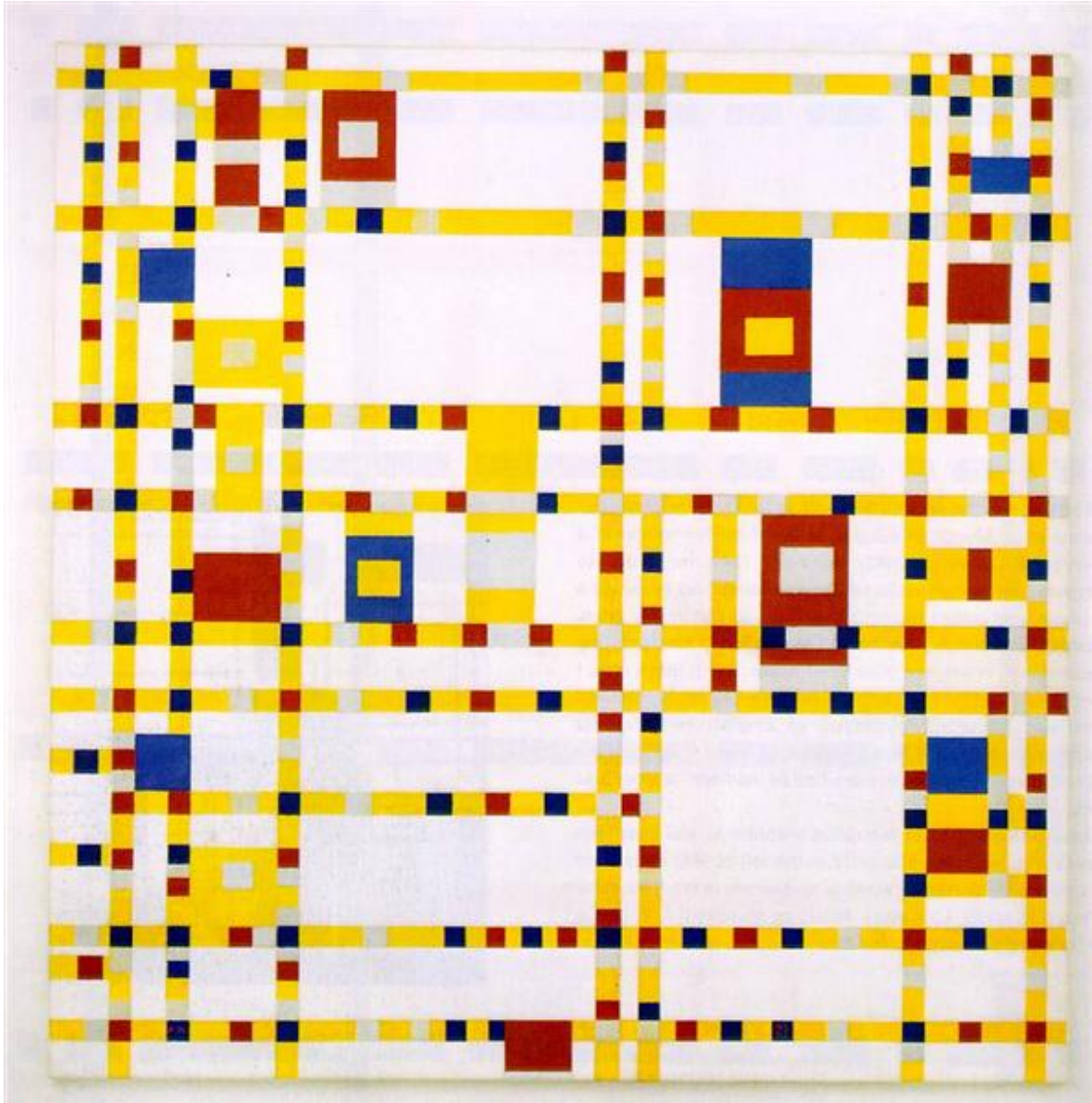


Figure 3c. *Broadway Boogie Woogie*, painting by Piet Mondrian (1942). The picture was presented to the participants with no caption.

According to specialists of Mondrian's work, Broadway Boogie-Woogie was painted after the two paintings reproduced below [Figures 3d and 3e]:

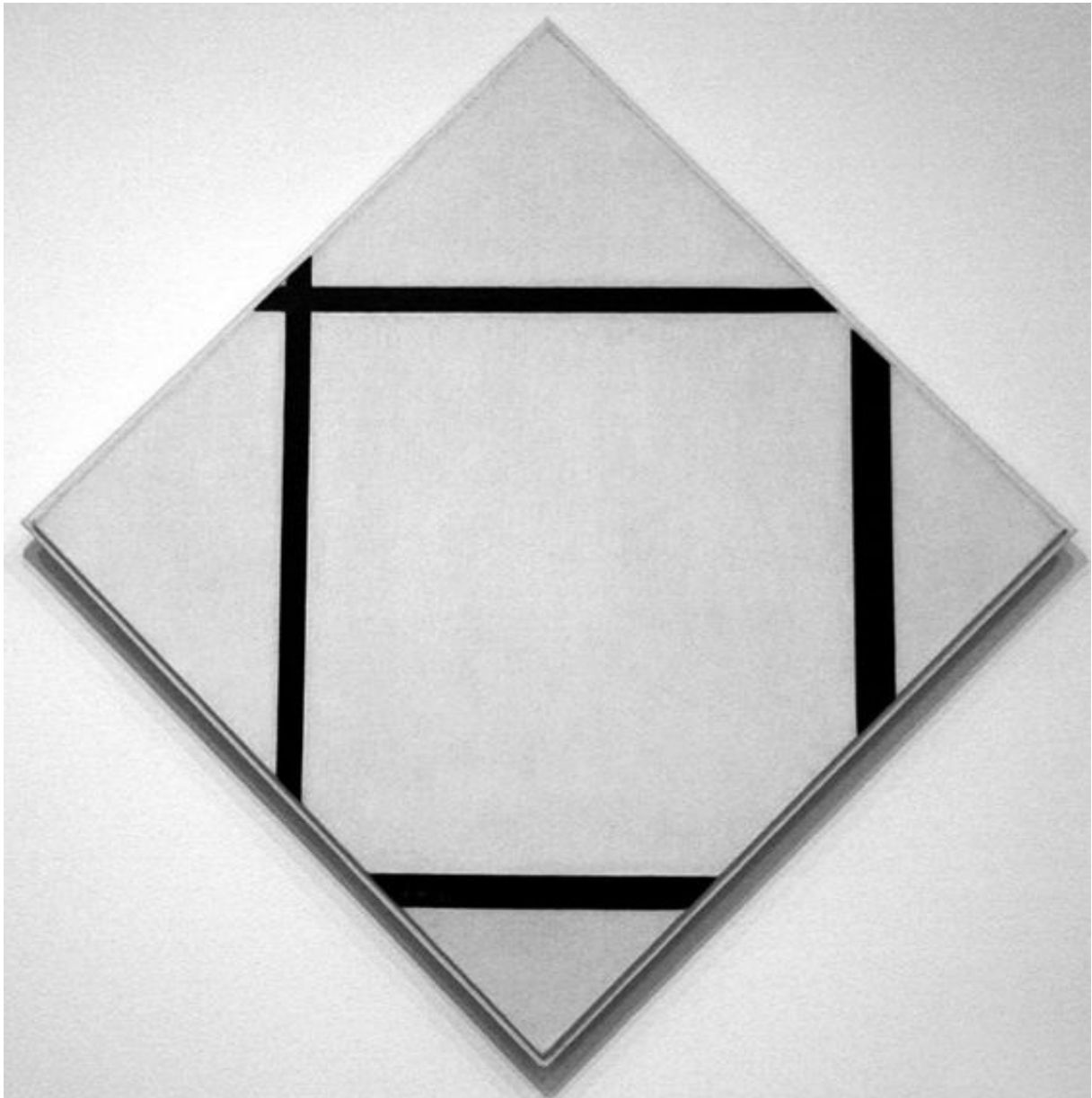


Figure 3d. *Tableau I: Lozange with Four Lines and Gray*, painting by Piet Mondrian, 1926. The picture was presented to the participants with no caption.

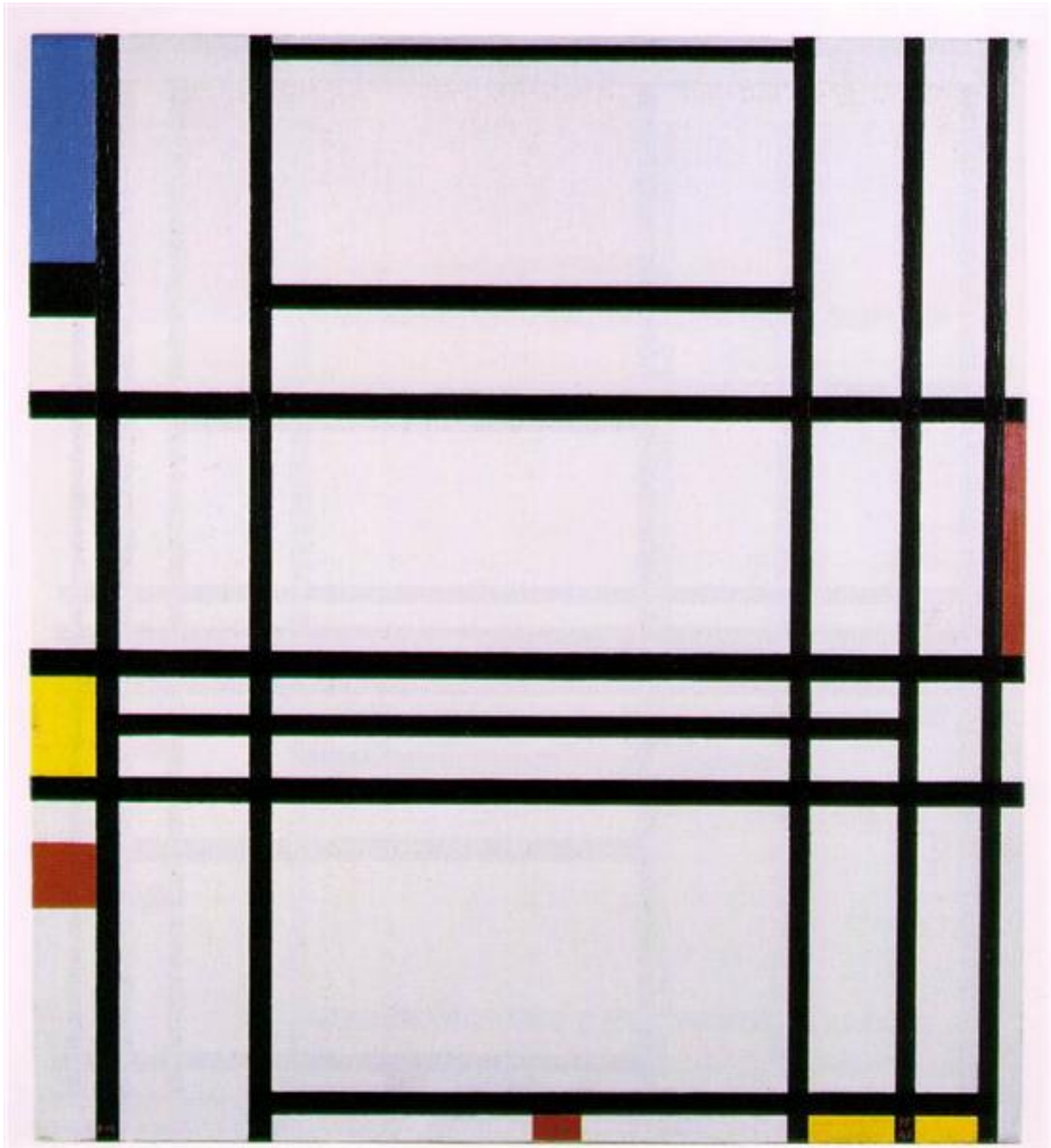


Figure 3e. *Composition n° 10*, painting by Piet Mondrian, 1939-1942. The picture was presented to the participants with no caption

After all the citizens have completed their evaluation, an art historian publishes information conclusively proving that Broadway Boogie-Woogie is not a work by Piet Mondrian, but a work by Gino Severini. Broadway Boogie-Woogie was actually painted after the two Severini paintings reproduced below [Figures 3f and 3g]:



Figure 3f. *Mare = Ballerina*, painting by Gino Severini, 1913. The picture was presented to the participants with no caption.



Figure 3g. *Dynamic Hieroglyphic of the Bal Tabarin*, painting by Gino Severini, 1912. The picture was presented to the participants with no caption.

Courtney and Darcy quickly agree about whether this new information matters for the evaluation of the originality and technical quality of Broadway Boogie-Woogie. But they start a dispute about the importance of this new information for the evaluation of the X of Broadway Boogie-Woogie: can we, or can we not, keep the marks given by the citizens in charge of evaluating the X of Broadway Boogie-Woogie?

Courtney thinks that we can't keep the marks given by the citizens in charge of evaluating X because considering Broadway Boogie-Woogie as a work of Mondrian or as a work of Severini can make a difference to its X.

Darcy thinks that we can keep the marks given by the citizens in charge of evaluating X because considering Broadway Boogie-Woogie as a work of Mondrian or as a work of Severini cannot make a difference to its X.

Generative Properties: Levinson's Perfect Synth

Courtney and Darcy work for a local media library. In order to make the catalog of this media library better known to the inhabitants of the neighborhood, Courtney and Darcy organize a

participative experiment. Randomly selected citizens are invited to listen to a set of records and to evaluate each recording according to one of the following three criteria: X, originality, and technical quality. Specifically, for each recording and for each of these criteria, the citizens assign a score from 1 to 10.

Among the recordings to be evaluated is a recording of Beethoven's Hammerklavier sonata for piano. Here is an excerpt of about ten seconds (please check that the sound on your device is active and then click on the Play button to listen to the excerpt): [the sound excerpt provided to the participants can be listened to [here](#)]²⁰.

After all the citizens have completed their evaluation, Courtney and Darcy realize that the liner notes of this recording do not mention the fact that the sonata was not played on a grand piano, but on a "HyperSound" synthesizer that perfectly imitates the sound of the piano, and is able to generate louder, higher, and lower sounds than a grand piano.

Courtney and Darcy quickly agree about whether this new information matters for the evaluation of the originality and technical quality of the recording. But they start a dispute about the importance of this new information for the evaluation of its X: can we, or can we not, keep the marks given by the citizens in charge of evaluating the X of this recording of the Hammerklavier Sonata?

Courtney thinks that we can't keep the marks given by the citizens in charge of evaluating the X of the recording because considering the Hammerklavier Sonata as played on a grand piano or as played on SuperSynth can make a difference to its X.

Darcy thinks that we can keep the marks given by the citizens in charge of evaluating the X of the recording because considering the Hammerklavier Sonata as played on a grand piano or as played on SuperSynth cannot make a difference to its X.

For each vignette, participants were asked with whom they agreed more, using the same rating scale as in the previous studies.

Participants and procedure

205 participants (age = 25.09; women: 130; men: 41; other or did not wish to report: 34) were recruited for this third study through the INSEAD-Sorbonne Université Behavioural Lab.

²⁰ The *Hammerklavier* excerpt (corresponding to mes. 239-248 of the 4th movement) was taken from Peter Bradley-Fulgoni's recording, available in open access through the IMSLP website ([https://imslp.org/wiki/Piano_Sonata_No.29,_Op.106_\(Beethoven,_Ludwig_van\)](https://imslp.org/wiki/Piano_Sonata_No.29,_Op.106_(Beethoven,_Ludwig_van))).

Participants recruited for the previous studies could not enroll in this third study. 97 participants were excluded for failing the attention check or the control questions, resulting in a final group of 108 participants (age = 24.67; women: 77; men: 29; other or did not wish to report: 2).

Sample size was determined a priori using G*Power 3.1. It recommended 132 participants for an ANOVA test with 6 groups and a large effect size (0.4). Given that we had to exclude all the participants enrolled in Studies 1 and 2, we could not quite reach that number.

Once the study was over, a lottery randomly selected 15 participants who each received 20 euros.

The study was conducted online using Qualtrics. Participants were first randomly assigned to one of two groups – corresponding to whether they would receive the “beauty” or the “intensity” version of the vignettes – and were then presented with our three vignettes in random order.

In short, our study followed a mixed factorial design, with *Aesthetic Property* (i.e., whether the question asked to the participants is about the evaluation of Beauty or Intensity) as a between-subjects factor and *Non-Perceptual Property* (i.e., whether the story pertains to Categorical, Proveniential, or Generative properties) as a within-subjects factor.

Results

First, using the same recoding procedure as in our previous studies, we found that a large majority of our participants sided with the empiricist position when it came to the evaluation of the artwork’s beauty (76.79% vs 23.21% for the anti-empiricist position). An exact binomial test showed that the difference between the majority and minority groups was significant ($p < 0.001$). A one sample t-test from the midpoint also showed that participants’ mean ratings were significantly higher ($M = 4.232$, $SD = 1.451$) than the threshold value of 3.5 ($t = 5.992$, $df = 55$, $p < 0.001$). This clearly replicates the results from our second study. However, intuitions were much more divided when it came to the evaluation of the artwork’s intensity: here, 59.62% sided with the empiricist position while 40.38% sided with the anti-empiricist position. And indeed, an exact binomial test showed that the difference between the majority and minority groups was not significant ($p = 0.212$). A one sample t-test from the midpoint also showed that participants’ mean ratings were not significantly different ($M = 3.724$, $SD = 1.560$) than the threshold value of 3.5 ($t = 1.789$, $df = 51$, $p = 0.07$). In other words, it seems that the relevance of the thought experiments discussed in the literature on aesthetic empiricism might indeed be largely dependent on the type of aesthetic properties targeted (i.e., expressive aesthetic properties such as intensity or more generic aesthetic properties such as beauty).

Second, in order to assess more precisely the impact of our two experimental factors (Aesthetic Property; Non-Perceptual Property) on participants' ratings (from 1 to 6, with the higher meaning closer to the empiricist position), the data were analyzed through a 2-way ANOVA, using the EZ package in R. Our statistical analysis revealed a main effect of Aesthetic Property ($df=106$, $F=8.403$, $p<0.005$, $\eta^2=0.029$) and a main effect of Non-Perceptual Property ($df=212$, $F=4.555$, $p=0.012$, $\eta^2=0.026$). No significant interaction between our two factors was found ($df=212$, $F=0.465$, $p=0.629$, $\eta^2=0.003$). A post-hoc Tukey test showed that ratings for the "Beauty" version of our vignettes were on average 0.508 higher than ratings for the "Intensity" version of our vignettes ($p=0.002$) (see Figure 4a; see also Figure 4b for rating distributions in each one of our conditions). Participants were thus significantly more inclined towards aesthetic empiricism when they had to assess the impact of non-perceptual properties on the artwork's beauty. This suggests that beauty might be generally taken to be particularly independent of non-perceptual properties. It also showed that ratings for the "Proveniential property" vignette (i.e., the Mondrian/Severini scenario) were 0.593 higher than ratings for the "Generative property" vignette (i.e., the piano/synthesizer scenario) ($p=0.010$). Similarly to what we found in Study 1, the artwork's historic-artistic context thus seems to have much less impact on the evaluation of its aesthetic properties (whether beauty or intensity) than the very way this artwork was generated.

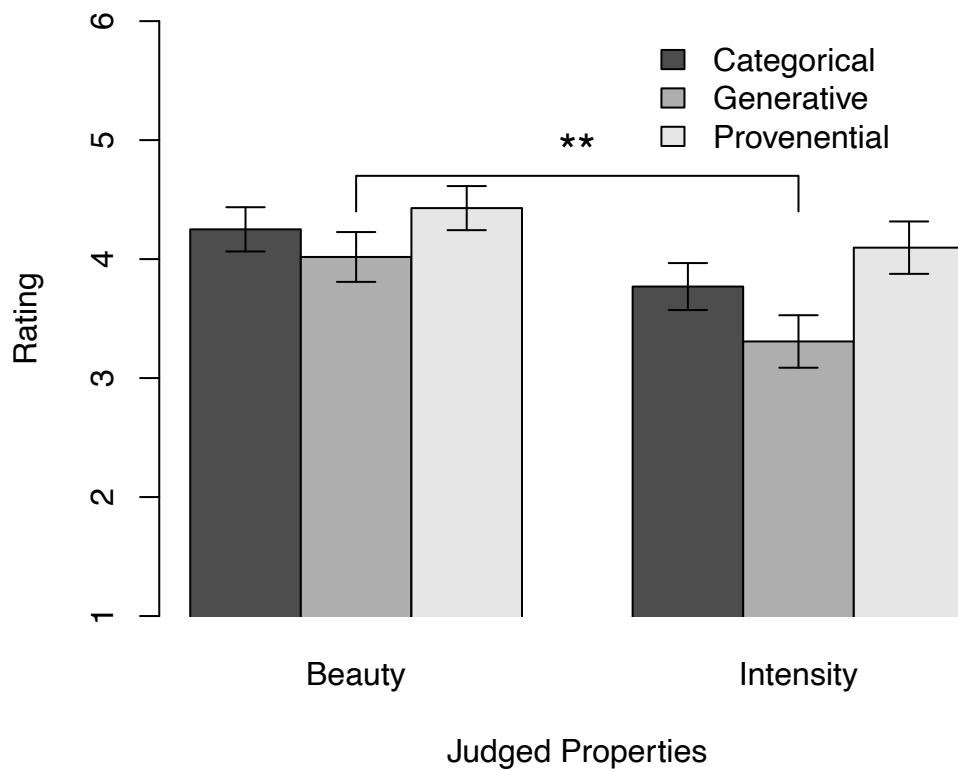


Figure 4a. Results from Study 3. Error bars show standard error (95% interval). * represents the significant outputs of the post-hoc tests, with a threshold of $p < .05$; **: $p < .01$; ***: $p < .001$

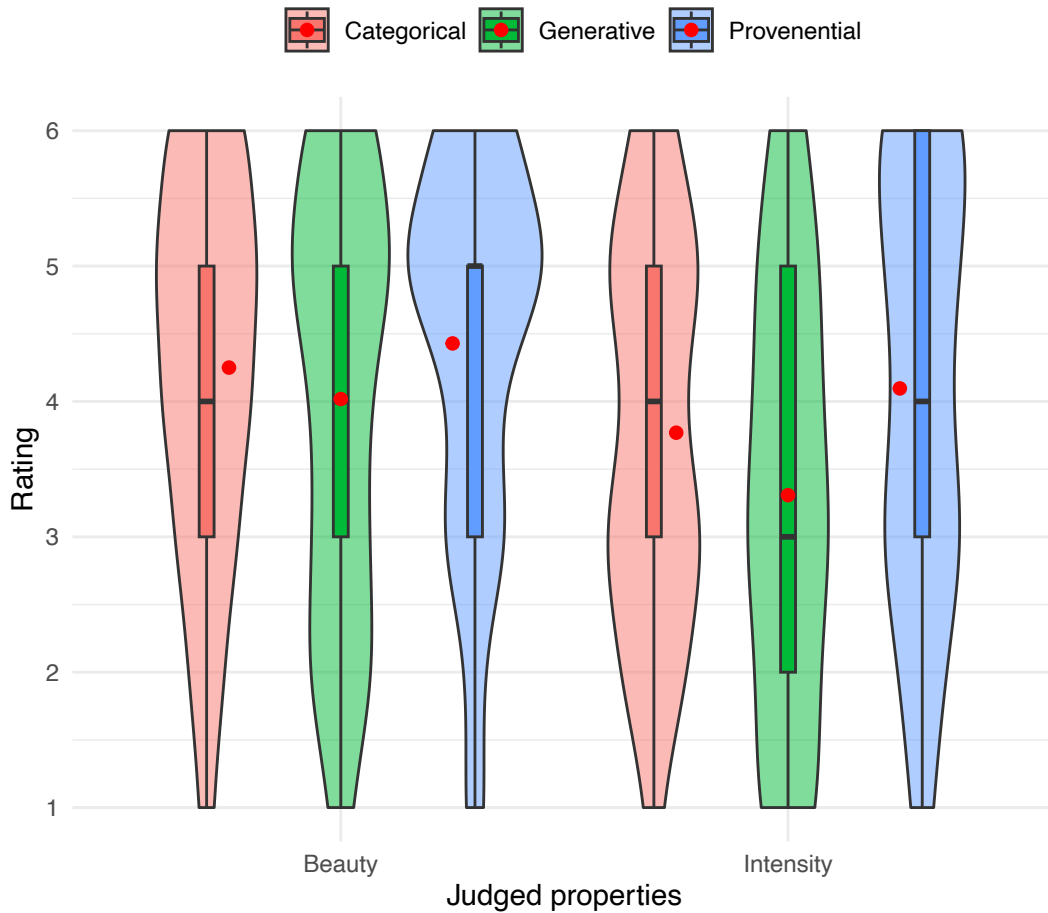


Figure 4b. Distribution of participants' ratings for Study 3.

Discussion

Whether or not an artwork's non-perceptual properties (such as its date of creation, the artistic category to which it is supposed to belong, or the hidden details of how it was conceived and produced) can have an impact on its aesthetic value is certainly amongst the most venerable debates in the philosophy of art. The three experiments reported above shed new light on this issue by showing how widely intuitions can vary on this topic, depending on the kind of aesthetic property that is targeted. On the one hand, most people did not believe that such non-perceptual properties can modulate our evaluation of an artwork's beauty – at least when “beauty” clearly refers to a certain type of aesthetic property rather than a catch-all term for the artwork's value *qua* artwork. Even when targeting aesthetic expressive properties more in line with the kind of thought experiments suggested by philosophers over the years – such as intensity – there was still no majority in favor of revising aesthetic judgments made about artworks after discovering new information about some of their non-perceptual properties

alone. On the other hand, our results also showed that artistic properties such as originality or technical achievement may be thoroughly impacted by a change in our beliefs about an artwork's non-perceptual properties such as proveniential properties, categorical properties, or generative properties. Taking those results together, two important conclusions might be drawn from our experiments for the ongoing philosophical discussion on aesthetic empiricism: first, the striking difference in participants' intuitions found between intrinsic aesthetic properties (beauty, intensity) and artistic properties (originality, technical achievement) provide additional ground for clearly differentiating *aesthetic* properties from the domain of *artistic* properties, even though the latter might also require an exercise of taste and may be included among the aesthetic properties or at least may be cited as reasons for aesthetic judgements; second, our results suggest that the philosophical thought experiments discussed in the present paper might be more justly considered to be relevant against artistic empiricism than against aesthetic empiricism: while the intuitions underlying *aesthetic* empiricism were significantly shared, *artistic* empiricism remained a minority view among our participants.

Our results should not either be interpreted as casting doubts on previous studies in the psychology of art appreciation which showed people to be strongly biased by non-perceptual information (such as the author's purported identity) when assessing artworks, including cardinal aesthetic properties such as its expressivity (Kroger & Margulis 2016). They are also compatible with results of experimental philosophy studies on moral character such as Doran (2021), which showed that participants were more likely to find a person's face beautiful if they were told that the person was morally good, thus suggesting that aesthetic judgments might be dependent on non-perceptual properties of a moral kind. Our own studies indeed investigated the participants' normative *beliefs*, not their actual behavior: people can very well think that we *should not* take into account anything else beyond an artwork's sensory surface when evaluating its beauty, while at the same time not being able to resist having their judgments influenced by all kinds of non-perceptual information. These two issues are largely independent and should be carefully distinguished.

A particularly striking aspect of our results is that not all aesthetic properties were equally taken by our participants to be sensitive to non-perceptual properties: they were much more inclined to believe that an artwork's aesthetic properties could depend on some of its non-perceptual properties when it came to the evaluation of its intensity rather than of its beauty. This suggests that aesthetic properties should not be treated in a monolithic way – which is in line with Carroll's suggestion (Carroll 1999) to distinguish between several, potentially overlapping, types of aesthetic properties, such as emotion properties (“somber”), character properties

(“bold”), Gestalt properties (“unified”), taste properties (“vulgar”) or reaction properties (“suspenseful”). In that perspective, the extent to which aesthetic properties seem to be sensitive to non-perceptive information could be used as a marker to delineate at least two broad categories of aesthetic properties: *expressive* aesthetic properties, including Carroll’s emotion and character properties, which may be construed as highly dependent on acts of imagination/interpretation (see e.g., Levinson 2006), and as such, are more likely to be sensitive to non-perceptual information about the artwork; and *formal* aesthetic properties, including Gestalt properties, which are more directly tied to perception, and as such, are less likely to be sensitive to non-perceptual information about the artwork.²¹ While “intensity” arguably belongs to the class of expressive aesthetic property, it is not clear whether “beauty” should be considered an example of formal aesthetic property. Our results suggest that it might generally be taken as such, or, at the very least, that it is indeed spontaneously construed as immediate, entirely residing in the artwork’s “sensuous appearance as it presents itself” to us, to quote Jerrold Levinson’s apt description of musical beauty (Levinson 2015: 59).

An alternative explanation for the difference we observed between “beauty” and “intensity” in our studies might relate to the semantics of such adjectives, rather than to the nature of the corresponding aesthetic properties. While “beauty” and “intensity” are both clearly gradable (an artwork can be more or less beautiful or intense), they may differ in being either relative gradables (i.e., tied to a contextually determined contrast class, like “big”) or absolute gradables (i.e., not tied to a contextually determined contrast class, like “spotted”). Here, “beauty” might be taken to be an absolute gradable whereas “intensity” would function as a relative gradable. However, recent experimental work on the topic has shown that the distinction between relative and absolute gradables does not clearly apply to aesthetic adjectives (Liao & Meskin 2017). If that is correct, then the prospects for explaining the difference we observed in semantical terms remains dim. It should be noted, however, that Liao & Meskin were careful to study both verdictive (“beautiful”) and substantive (“elegant”) aesthetic adjectives, but did not consider expressive adjectives such as the ones employed in the thought experiments by Gombrich, Walton, and Levinson. So the hypothesis that expressive aesthetical terms, unlike formal ones, do pass the tests for relative gradability remains to be tested and may be the topic of further experimental work.

²¹ Carroll classified beauty, along with the sublime, among the reaction properties, plausibly because we react to beauty with pleasure and to the sublime with awe. This classification is compatible with the view that some beauties at least will count as formal, if we take our pleasure from the Gestalt of the work, and some sublimities as expressive, if awe is triggered by some intense expressive properties.

A third option would be that “beauty” is a thin aesthetic concept, i.e., a concept that conveys only an evaluative meaning with no or little descriptive content, while “intensity” is rather a thick aesthetic concept, i.e., a concept that conveys at the same time a nontrivial description and an evaluation (see Bonzon 2009). An interesting experimental follow-up to the studies reported here would thus be to see if aesthetic properties that are both formal and denoted by thick concepts – such as “elegance” – would lead to the same results as “beauty” or to a pattern more similar to what we observed in the case of “intensity”. Such a study may disentangle expressive aesthetic properties from thick aesthetic concepts and clarify what makes certain aesthetic properties more reluctant to anti-AE intuitions.

Our studies also invite considering more carefully how anti-AE intuitions might be more or less powerful depending on the artistic medium. The striking differences observed between Study 1 and Study 2 might also hint at how beauty is cashed out differently in music and in painting. Alternatively, it might well be the case that generative properties are taken to play a more important role in evaluating music’s beauty because of its status as a performing art.

Of course, all the above discussion is dependent on accepting that our experiments can actually shed light on the philosophical debate surrounding aesthetic empiricism. Anti-empiricists could still raise doubts about the validity of our experiments for the issue at stake by pointing out that our participants were far from being expert art appreciators (of the sort artists, art critics, or philosophers of art can be), and that could have been enough to make them insensitive to the arguably subtle effects brought forward in anti-empiricist thought experiments. In fact, as pointed out above, many critics of AE acknowledge that it is a commonsensical view, or at least a “seemingly” commonsensical one (Walton 1970: 367). From this point of view, it may be argued that our results are compatible with the view that AE is commonsensical but false. It just turns out that naïve subjects lack some part of the required expertise that is responsible for the aesthetic intuitions embedded in the indiscernibility arguments of Gombrich, Walton, and Levinson.

By rejecting inferences about the reliability of philosophers’ intuitions based on data about so-called naïve subjects, such a response is committed to what is known in the experimental philosophy literature as an “expertise defense” (Williamson 2011). As a general response to experimental findings, the expertise defense remains controversial (Machery 2017: section 5.3). But it does not mean that a restricted version, specific to the aesthetic domain is not available. In fact, Weinberg (2019) suggested that the prospects for an *aesthetic* expertise defense are significantly brighter than for the general expertise defense. His suggestion is based on two points. First, the intuitions typically featured in aesthetic thought experiments are consensual

and reinforced by familiarity with similar cases in the philosophy of art. To substantiate this point, Weinberg refers to Danto's gallery of indiscernibles where the classification as art of the monochrome made by the fictional character named "J" is backed by the resemblance between J and the historical Jasper Johns, assuming on the part of the reader familiarity with Johns' work and similar work of the same period by Warhol and others. However, the cases studied here are importantly different in that the crucial aesthetic judgments depended on a highly improbable counterfactual assumption, i.e., the existence of the GUERNICA category or the event that Severini was the real author of *Broadway Boogie-Woogie*, for which obvious counterparts in the history art are hard to find. So it is impossible to rely on an art-historical consensus here. Second, Weinberg hypothesizes that the epistemically deleterious sensitivity of naïve subjective to extraneous factors (such as order of presentation, socio-economic background, or personality traits) documented at length in the experimental-philosophical literature should be reduced by presenting actual artworks as opposed to filled-in verbal descriptions. However, our Study 3 showed that substituting verbal descriptions of fictional works with actual works did not affect the distribution of aesthetic intuitions. So, either Weinberg's hypothesis is false, or the pro-AE pattern observed in our Study 3 in the case of beauty cannot be explained by extraneous factors. Although we could not control for all possible such factors, we did control for order effects and usualness effects. This naturally leaves open the possibility that our pro-AE results about beauty may be explained by other extraneous factors. In particular, in light of extended work in the sociology of taste following the foundational work of Bourdieu (1979), it is not implausible that the endorsement of AE varies with some socio-economic factors. Again, this is as much a limitation of the present study as an avenue for future empirical research. In any case, if a plausible aesthetic expertise defense is to be deployed in response to our results, it needs to be different than the one sketched by Weinberg.

There are also more general reasons to be skeptical of an aesthetic expertise defense that would grant art historians, artists, and philosophers of art a distinctive authority on aesthetic concepts and judgments. Many aspects of our everyday lives are indeed infused with aesthetic appreciation and judgment (Lopes, Nanay & Riggle 2022), well beyond the relation each one of us may or may not have to traditional artworks. There are thus no reasons to think that our participants lacked the mastery of the relevant aesthetic concepts to the extent that they were unable to have reliable intuitions on aesthetic matters. That being said, it is true that our experimental vignettes manipulated subtle differences in background information (e.g., whether Beethoven's *Hammerklavier* piano sonata is performed on a piano or a HyperSound synthesizer). Even though we tried to be as explicit as possible when introducing our target

contextual information, it remains possible that the relevance of such information could only have been accessible to people exhibiting a sufficient familiarity with the artistic practice under consideration (e.g., people who know in detail how a piano actually works). Further experiments on the topic could thus control more precisely for artistic expertise, to see whether anti-empiricist intuitions are more likely to arise amongst participants with a high familiarity with the artistic domains discussed throughout the experimental vignettes.

All these limitations notwithstanding, our studies show that the central intuitions elicited by the classical indiscernibility arguments popularized by Gombrich, Walton, and Levinson depend on the class of aesthetic properties (expressive aesthetic properties vs formal aesthetic properties; aesthetic properties vs artistic properties – to the extent that the latter are aesthetic properties at all) and are thus not sufficiently robust to refute conclusively aesthetic empiricism. We do not mean to suggest that aesthetic empiricism should, as a result, be reinstated as the prevailing doxa, as it arguably was in mid-twentieth-century Anglo-American aesthetics. However, it appears that aesthetic empiricism, or some suitably qualified version thereof, clearly distinguished from artistic empiricism, still deserves to be taken seriously, or at least criticized anew.

References

Alperson, Philip, 1984, On musical improvisation, *Journal of Aesthetics and Art Criticism*, 43, pp. 17–29.

Beardsley, Monroe C. 1958. *Aesthetics: Problems in the philosophy of criticism*. New-York & Burlingame: Harcourt, Brace & World, Inc.

Bell, Clive. (1914). *Art*. New-York, Frederick A. Stokes Company.

Bonzon, Roman (2009). Thick aesthetic concepts. *The journal of aesthetics and art criticism*, 67(2), 191-199.

Bourdieu, Pierre, 1979, *La distinction*. Paris, Éditions de Minuit.

Brown, Lee B., 1996, Musical Works, Improvisation, and the Principle of Continuity, *Journal of Aesthetics and Art Criticism*, 54:4, pp. 353–369.

Canonne, Clément (2018). Listening to Improvisation, *Empirical Musicology Review*, vol. 13, no. 1, p. 1–14.

Carroll, Noël, 1999, *Philosophy of Art. A contemporary introduction*. London and New York, Routledge.

Currie, Gregory, 1989, *An Ontology of Art*, London, Palgrave Macmillan.

Danto, Arthur C., 1981, *The Transfiguration of the Commonplace*, Cambridge MA, Harvard University Press.

Danto, Arthur C. 2003. *The Abuse of Beauty: Aesthetics and the Concept of Art*. Open Court Publishing.

Davies David, 2003. *Art as Performance*. Malden MA, Blackwell.

Davies, David, 2009. The Artistic Relevance Of Creativity. In Michael Krausz, Karen Bardsley & Denis Dutton (eds.), *The Idea of Creativity* (pp. 213-233). Leiden. Brill.

Davies, David, 2011. *Philosophy of the Performing Arts*. Chichester, Wiley-Blackwell.

Davies, Stephen 2001, *Musical works and performances. A philosophical exploration*. Oxford, Oxford University Press.

Dodd, Julian, 2007. Sounds, instruments and works of music. In Kathleen Stock (ed.), *Philosophers on Music: Experience, Meaning, and Work*. Oxford University Press.

Doran, Ryan P. (2021). Moral beauty, inside and out. *Australasian Journal of Philosophy*, 99(2), 396-414.

Ellington, Duke (1973). *Music is my mistress*. New-York: Da Capo Press.

Gombrich, Ernst H., 1961, *Art and illusion*, London, Phaidon Press.

Gooley, Dana (2018). *Fantasies of improvisation: Free playing in nineteenth-century music*. Oxford: Oxford University Press.

Grove, George, 1903. Mendelssohn's Overture to 'A Midsummer Night's Dream'. *The Musical Times and Singing Class Circular*, 44(729), 728-738.

Hamilton, Andy, 2000, The art of improvisation and the aesthetics of imperfection, *British Journal of Aesthetics*, vol. 40, no. 1, pp. 168–185.

Hanson, Louise, 2013. The reality of (non-aesthetic) artistic value, *The Philosophical Quarterly*, vol. 63, no. 252.

Huddleston, Andrew, 2012, In defense of artistic value, *The Philosophical Quarterly*, vol. 62, no. 249, p. 705–714.

Huron, David. 2008. *Sweet anticipation: Music and the psychology of expectation*. Cambridge: MIT press.

Iseminger, Gary, 2010. Sonicism and jazz improvisation, *Journal of Aesthetics and Art Criticism*, vol. 68, n. 3, pp. 297–299.

Kania, Andrew, 2008. Piece for the end of time: In defence of musical ontology. *The British Journal of Aesthetics*, 48(1), 65-79.

Kroger, Carolyn, & Margulis, Elizabeth H. (2017). "But they told me it was professional": Extrinsic factors in the evaluation of musical performance. *Psychology of Music*, 45(1), 49-64.

Kruger, Justin, Wirtz, Derrick, Van Boven, Leaf, & Altermatt, T. William (2004). The effort heuristic. *Journal of Experimental Social Psychology*, 40(1), 91-98.

Kulka, Tomas (1981). The artistic and the aesthetic value of art. *The British Journal of Aesthetics*, 21(4), 336-350.

Lamarque, Peter (2010). "Aesthetic Empiricism", in *Work and Object*, Oxford, Oxford University Press, 122—138.

Levinson, Jerrold. (1980). What a musical work is. *The journal of philosophy*, 77(1), 5-28.

Levinson, Jerrold (2006). "Musical Expressiveness as Hearability-as-Expression", in *Contemplating art: Essays in aesthetics*. Oxford, Oxford University Press, pp. 91-108.

Levinson, Jerrold (2011). "Artworks and the Future", in *Music, Art, and Metaphysics*, Oxford, Oxford University Press, pp. 179-214.

Levinson, Jerrold (2015). "Musical Beauty", in *Musical concerns: Essays in philosophy of music*. Oxford, Oxford University Press, pp. 58-66.

Levinson, Jerrold (2017). "Aesthetic contextualism", in *Aesthetic pursuits: Essays in philosophy of art*. Oxford, Oxford University Press, pp. 17-27.

Liao, Shen-Yi, & Meskin, Aaron (2017). Aesthetic adjectives: Experimental semantics and context-sensitivity. *Philosophy and Phenomenological Research*, 94(2), 371-398.

Lopes, Dominic McIver (2011). The myth of (non-aesthetic) artistic value, *The Philosophical Quarterly*, Vol. 61 No. 244.

Lopes, Dominic McIver., Nanay, Bence., & Riggle, Nick. (2022). *Aesthetic Life and why it Matters*. Oxford University Press.

Machery, Édouard (2017). *Philosophy within its proper bounds*, New York, Oxford University Press.

Nanay, Bence (2015). Cognitive penetration and the gallery of indiscernibles, *Frontiers in Psychology*, Volume 5 - 2014 | <https://doi.org/10.3389/fpsyg.2014.01527>

Prall, David W. (1936). *Aesthetic analysis*. New-York, Thomas Y. Crowell.

Puy, Nemesio Garcia-Carril (2021). Methodological worries on recent experimental philosophy of music, *Philosophical Psychology*, volume 35, issue 3.

Sibley, Frank, (1959). Aesthetic Concepts, *The Philosophical Review*, Vol. 68, No. 4, pp. 421-450.

Sibley, Frank (1965). Aesthetic and Non-Aesthetic, *The Philosophical Review*, Vol. 74, No. 2, pp. 135-159.

Stecker, Robert (2012). Artistic value defended. *The Journal of Aesthetics and Art Criticism*, 70(4), 355-362.

Stokes, Dustin (2014). Cognitive Penetration and the Perception of Art, *Dialectica*, vol. 68, No. 1, pp. 1-34. DOI: 10.1111/1746-8361.12049

Stokes, Dustin (2018). “Rich Perceptual Content and Aesthetic Properties”, in Anna Bergqvist & Robert Cowan (eds.) *Evaluative Perception*, Oxford: Oxford University Press, pp. 19-41.

Walton, Kendall (1970). Categories of Art, *The Philosophical Review*, vol. 79, n.3, pp. 334–367.

Weinberg, Jonathan (2019). “Are Aestheticians’ Intuitions Sitting Pretty?”, in Florian Cova & Sébastien Réhault (eds), *Advances in experimental philosophy of aesthetics*, New-York, Bloomsbury, pp. 267-288.

Williamson, Timothy (2011). Philosophical expertise and the burden of proof. *Metaphilosophy*, 42, 215–29.

Wollheim, Richard, 1980. Art and its objects: with six supplementary essays. New York: Cambridge University Press. Edited by Richard Thomas Eldridge.

Zangwill, Nick, 1999. *The Metaphysics of Beauty*, Ithaca & New York, Cornell University Press.