

The Role of Teleological Thinking in Judgments of Persistence of Musical Works

Elzė Sigutė Mikalonytė

Vilnius University, Institute of Philosophy

Vilius Dranseika

Vilnius University, Institute of Philosophy

Jagiellonian University, Institute of Philosophy and Interdisciplinary Centre for Ethics

Abstract. In his paper “The Ontology of Musical Versions: Introducing the Hypothesis of Nested Types” (2019), Nemesio Puy raises a hypothesis that continuity of the purpose is both a necessary and a sufficient condition for musical work’s identity. Puy’s hypothesis is relevant to two topics in cognitive psychology and experimental philosophy. The first topic is the prevalence of teleological reasoning about various objects and its influence on persistence and categorization judgments. The second one is the importance of an artist’s intention in the categorization of artworks. We tested the teleological hypothesis across three studies. Vignettes in these three studies describe a musical work being changed in some of these aspects: (a) purpose either changed or retained; (b) score either changed or retained; (c) change is made either by the same or a different composer. The results suggest that teleological considerations impact judgments on the persistence of musical works, but this impact appears to be relatively weak. The results also suggest that persistence judgments strongly depend on whether acoustical properties were changed, while whether the change was made by the original composer seems to be relatively unimportant.

Keywords. Ontology of musical works, teleology, experimental philosophy.

1. Introduction

In his paper “The Ontology of Musical Versions: Introducing the Hypothesis of Nested Types” (2019), Nemesio Puy aims to formulate a descriptive theory of musical ontology which explains the nature of different versions of the same musical works. He modifies type/token theory (according to which, a musical work is an abstract entity instantiated by its performances) and claims that each musical work can have more than one sound structure. Instead of one, there may be two or more sound structures that are similar enough to be versions of the same musical work (Puy 2019, 251). Thus, Puy modifies the two-level type/token theory into one of many levels,

meaning that one musical work can have many different sound structures which share the same purpose intended by the composer, and each of those sound structures can be instantiated by musical performances. The author aims to “analyze *the standard view* on versions embodied in our musical practices” (Puy 2019, 242) when he stresses the crucial role of the shared purpose in joining different sound structures into one musical work.

Therefore, when seeking to create a theory accommodating our intuitions, Puy raises a hypothesis about the way we ordinarily think about the persistence of musical works. He claims that “different sound structures of a work’s versions must keep the overall point or purpose of the work” for us to consider them as two versions of the same musical work (Puy 2019, 243). According to Puy, the way we ordinarily think about musical works’ identity depends, most importantly, on their perceived purposes. As long as the original purpose is retained, continued identity of a musical work is compatible with changes to its sound structures and instrumentations (Puy 2019, 244). Puy’s hypothesis is relevant to two important topics in cognitive psychology and experimental philosophy. The first topic is the prevalence of teleological reasoning about various objects and its influence on judgments about persistence and categorization, while the second topic is the importance of the artist and their intention in the categorization of artworks, surpassing the significance of an object’s appearance.

1.1. Teleological reasoning. The literature of cognitive psychology suggests that teleology is the basic mode of ordinary thinking. Teleological intuitions about both natural objects and artifacts, a tendency of seeing a purpose instead of thinking in physical-causal terms, is considered to be a cognitive default (Kelemen, Rottman, and Seston 2013). Teleological explanations of why various objects exist, in other words, explanations in terms of their purpose, are not only inherent in the reasoning of children (DiYanni and Kelemen 2005, Kelemen 1999), but also under certain conditions (e.g., when making fast speeded judgments) in reasoning of adults (Kelemen and Rosset 2009) or even adult scientists (Kelemen, Rottman, and Seston 2013).

Since many contemporary metaphysicians aim to formulate theories that respect folk intuitions, it is important to take into consideration that the folk tend to think in the teleological mode. One example of folk teleology that is relevant to philosophical problems is that the folk operate under a teleological view of persistence. It has been noticed by David Hume, who wrote that:

There is, however, another artifice, by which we may induce the imagination to advance a step farther; and that is, by producing a reference of the parts to each other, and a combination to some common end or purpose. A ship, of which a considerable part has been changed by frequent reparations, is still considered as the same; nor does the

difference of the materials hinder us from ascribing an identity to it. The common end, in which the parts conspire, is the same under all their variations, and affords an easy transition of the imagination from one situation of the body to another (2007, 167-168).

In folk theory, the persistence of ordinary material objects is held to be dependent on their function preservation. A material object is thought to survive extensive change as long as its function remains the same (Rose 2015). Research in folk mereology also suggests that a collection of things forms a new single composite object when a composite object has a new function and a new purpose (Rose and Schaffer 2017).

Moreover, Rose and Nichols propose that teleological essentialism generates category judgments (2020). The prevalent view posits that the categories are shaped around what is called an “essence placeholder”. There is an intuitive folk belief that drives categorization judgments and according to it, all objects have some kind of unspecified essence (Medin and Ortony 1989). In contrast, Rose and Nichols argue that the “essence placeholder” is actually filled by the object’s telos. Their research suggests that telos is considered as the object’s essence, and that the object is categorized based on either change or preservation of it (Rose and Nichols 2019). Perceptual properties do not seem to be as important in the process of categorization. However, they notice that it is still unclear how much teleology and other features of an object contribute to categorization judgments concerning *artifacts* (Rose and Nichols 2020).

The latter two studies may raise some questions of relevance to the domain of art, because research in cognitive psychology suggests that psychological essentialism is inherent to the way people ordinarily think about works of art (Rabb, Brownell, and Winner 2018; Newman 2016).

1.2. Conceptual and perceptual factors in categorization and valuation of artworks.

There are two opposing views on the importance of perceptual and non-perceptual properties in categorization and valuation of musical works in philosophy of art. Aesthetic contextualism claims that works of art have no clear identity and no aesthetic properties independent from the context of their creation (Levinson 2017, 20). The second view, aesthetic empiricism, claims that the work’s aesthetic properties are present in its acoustic appearance, that is why the work’s appearance is enough to provide identity conditions and to give rise to aesthetic properties regardless of the context of creation (Dodd 2007, 6).

Aesthetic empiricism is not a widely held position. According to Dutton (2009), people think about artworks not only in terms of their perceptual features, but also in terms of their being results of artists’ creative performances. Research in psychology shows that not only perceptual properties, but also conceptual factors are considered while assessing the artworks. Information about the artist’s mental state and their physical connection with the artwork is very important in

the categorization and valuation of an artwork (Newman and Smith 2018). This kind of research supports the hypothesis that artworks are held to be mental and physical extensions of their creators.

One of the aspects of psychological essentialism regarding artworks is a belief in contagion, or a belief that through physical contact with the creator, artworks can be “contaminated” with her essence, and, more broadly, that artworks embody the essence of people who created them. Research on intuitions about visual artworks shows that people see original artworks as more valuable than identical duplicates, and this effect is explained by tendency to assess the object as a unique creative performance and by the physical contact with the artist, which “contaminates” the object with her essence (Newman and Bloom 2012).

Psychological essentialism is not only important in the valuation of art; it is also held to be responsible for judgments of persistence and category membership. A belief that a creator extends her essence into her artwork is also relevant to judgements on when an artwork stays or ceases to be “the same”. Bloom claims that like all artifacts, artworks are also categorized as belonging to some specific category in case its creator intended for it to belong to that category. Creator’s intentions are more important in categorization process than the object’s appearance or function (Bloom 1996). People see original visual artworks not only as more valuable than identical duplicates, but they also do not consider the latter to be “the same” artwork (Newman, Bloom 2012). Moreover, the duplicate artworks made by the same individual are more likely to be seen as continuers of the original than those made by a different person. Judgements on the identity of artworks resemble those on personal identity (in both cases, sameness of the physical material is held to be important for persistence), again supporting the hypothesis that art objects are seen as extensions of their creators (Newman, Bartels, and Smith 2014).

1.3. Visual and aural art. A big part of empirical research on the importance of non-perceptual factors in categorization and evaluation of artworks has concentrated on visual artworks, and on what Goodman calls autographic artworks (where the distinction between original and forgery is significant, as opposed to the allographic artworks – those that do not allow forgery (1968, 113)). Allographic artworks might have a different relation to their creator since they cannot be identified with physical objects. The importance of the physical continuity between the artwork and its creator, as well as the importance of the continuity of the physical material (Newman, Bartels and Smith (2014)), cannot be as significant in the case of allographic works as they are in the case of autographic works. It is widely believed that authentic artworks can be distinguished from forgeries (that are not “the same” artworks) by their history and connection to the creator. But can these factors also explain judgments on when a musical work’s versions are of “the same” musical work? Research suggests that intentional creation of a musical work can

also be perceived to “contaminate” the work with the creator’s essence: there is one study showing that valuation of music depends on composer’s morality even if in this situation there is no physical object that could be “contaminated” (Stavrova et al. 2016). However, paintings are normally taken as paradigmatic examples of artworks in the psychology of art. Thus, we know much less about the non-visual and allographic arts and whether the existing findings pertain to them.

No research on the role of teleological reasoning in individuation judgments exists in the ontology of musical works. However, there are several benefits that this kind of research could bring. First, and most importantly, it can inform ontologists of music. Second, it can also put to test the hypothesis of teleological essentialism in the ontology of art. Third, it can help to extend research in psychological essentialism about artworks to the aural and allographic types of works.

1.4. Implications for the ontology of musical works. Puy’s teleological hypothesis might be contrasted with Dodd’s claim that any change in work’s sound structure results in the creation of a new work. According to Dodd, it is not possible to change the work; what is called “a new version” is, in fact, a separate work, and the earlier and later versions of it are two distinct works (2007, 88). This position follows from the theory of individuation called sonicism, which tells that works are numerically identical in case they are acoustically indistinguishable. Dodd sees sonicism as intuitive and compliant with our aesthetic practice (*ibid.*, 6), thus our study could be seen as an effort to examine which of these positions – either of Puy or of Dodd – is more intuitive.

The most important question for philosophers in this context is whether they should incorporate teleology into their theories. Some philosophers see folk teleology as a reason for dismissing folk intuitions and giving up on descriptivist metaphysical theories (Rose and Schaffer 2017, Rose 2015). In the ontology of art, this kind of decision does not seem obvious, especially because it deals with artifacts, which have traditionally been seen as intention-dependent objects. Puy’s hypothesis and his descriptive theory, which aim to incorporate teleological reasoning about musical works, serve as a clear example that not all ontologists of art will see teleology as a reason to dismiss folk intuitions.

There is no consensus on whether teleological intuitions should be integrated into theories of artifactual kinds. Psychological essentialism applies to both natural or artifactual kinds (Gelman 2013) and the experiments by Rose and his colleagues are not limited to examples with natural kinds, but they also include artifacts (Rose and Nichols 2020). The authors draw no clear line between these two kinds of objects in their proposal to ignore folk intuitions. Dink and Rips, on the other hand, see teleological explanations of artifacts as potentially appropriate and the least problematic compared to other kinds of objects (Dink and Rips 2017).

Philosophers who aim to integrate ontology of art into general metaphysics may want to disregard teleological intuitions regardless of whether an object belongs to an artifactual or to a

natural kind. On the other hand, for philosophers of art who see no problem in creating ontological categories that pertain only to art objects, teleological intuitions can be accepted into theories of work identity.¹

1.5. Research objectives. Our main aim is to test the teleological hypothesis formulated by Puy across three studies. Purposes, however, come in different flavors. According to Puy (2019, 244), the purpose of a work may be an expressive one, a descriptive one (the musical description of a place, a character, or a process), or a purely musical one. For our studies, we created vignettes reflecting different types of purposes. According to a strong version of the teleological view, as it is formulated by Puy, continuity of the purpose is both a necessary and a sufficient condition for work's identity. We may also formulate a weak version of teleological view: it would argue that study participants will be *more likely* to believe that there are two distinct musical works if the purpose of a musical work is changed, and more likely to believe that there is one work if the purpose is retained. Sonicism, on the other hand, predicts that changes in purpose, if they happen without any changes in acoustic properties, will not influence judgments on work identity.

Although our main goal is to check whether changing the purpose has an influence on persistence judgments, we also manipulate two other aspects – changes in the score and in composer. Changes in the score inevitably result in changes in the way musical works sound, therefore, this variable of our vignettes is especially relevant for sonicism in musical ontology. Manipulating changes in score allows us to examine the importance of perceptual features of musical works. Both teleological and sonicist hypotheses agree that changes in acoustic properties tend to result in the creation of a distinct musical work. However, teleological hypothesis claims that changes to the score may be considered insignificant as long as the overall purpose of the work is retained, while sonicism claims that two even marginally different scores are scores of two distinct works (Dodd 2007, 90). For Puy, structural differences are secondary. He claims that the purpose is what constrains the extent of the structural differences that different versions may exhibit while still being considered the same musical work (Puy 2019, 243). If a strong version of sonicist hypothesis is correct, participants will treat continuity of the score as a necessary and sufficient condition for identity; if a weak version is correct, continuity of the score will have an influence on individuation judgments, but will not be decisive.

Manipulating whether the changes were made by the original or another composer also allows us to take a look at the hypothesis that our social practices allow musical works being changed only by their original creators (Friedel 2020). If a strong version of this hypothesis is correct, participants will treat continuity of the author as a necessary condition for work identity. By contrast, Puy claims that work's versions can be made by different composers (Puy 2019, 243). If a weak version of Friedel's hypothesis is correct, study participants will be more likely to believe

that a change made by a different composer results in the creation of a distinct work than one made by the original composer.

Accordingly, strong and weak versions of the three hypotheses can be summarized as follows:

[*Strong teleological hypothesis*] Study participants will believe that there is one and the same musical work if the purpose is retained, and they will believe that there are two distinct musical works if the purpose is changed.

[*Weak teleological hypothesis*] Participants will be more inclined to believe that there is one musical work if the purpose is retained, and two distinct works if the purpose is changed.

[*Strong sonicist hypothesis*] Participants will believe that there is one musical work if the score is retained, and two distinct works if the score is changed.

[*Weak sonicist hypothesis*] Participants will be more inclined to believe that there is one musical work if the score is retained, and two distinct works if the score is changed.

[*Strong original creator's hypothesis*] Participants will believe that there is one musical work if the change is made by the same composer, and two distinct works if the change is made by a different composer.

[*Weak original creator's hypothesis*] Participants will be more inclined to believe that there is one musical work if the change is made by the same composer, and two distinct works if the change is made by a different composer.²

As our main aim in this paper is to test the teleological hypothesis, vignettes in our studies describe a musical work with its (a) purpose either changed or retained. However, we also vary – in a less systematic manner in Studies 1 and 2 and systematically in Study 3 – whether (b) the score is changed or retained; and whether (c) the change is made by the same or a different composer. This allows us to compare the teleological hypothesis to the other two – the sonicist and the original creator's hypotheses.

2. Study 1

Six separate studies that share the same basic design were conducted to evaluate the impact of retaining or changing the purpose on judgments of individuation of musical works.

Participants. Participants in Studies 1a-f were recruited on Prolific.ac to take part in this online study. Participants were US or UK nationals who indicated English as their first language. See Table 1 for information on participants in studies 1a-f.

Study	N	<i>M</i> _{age} (SD); age range	% female/male/non-binary
1a	149	37.0(13.2); 18-68	59%/40%/1%
1b	134	34.9(13.3); 18-69	66%/31%/3%
1c	223	35.8(12.1); 18-75	64%/35%/0%
1d	220	34.3(11.1); 18-72	68%/30%/2%
1e	201	36.1(13.0); 18-75	58%/39%/3%
1f	228	34.7(13.7); 18-76	62%/38%/0%

Table 1. Sample characteristics for studies 1a-f. Percentages may not add up to 100% due to rounding.

Methods and Materials. Each participant received one out of two vignettes: either with purpose retained or purpose changed. In addition, we also manipulated changes to the score and changes of the composer across types of vignettes. Thus, the score was described as either changed (in half of the studies – 1a, 1b, 1e) or retained (1c, 1d, 1f), and the change as being made either by the original (1c, 1e, 1f) or a second composer (1a, 1b, 1d). After reading a vignette, participants were asked to imagine two concerts in which two versions of the score are used, and to answer the question: “Were the listeners in these two concerts listening to two performances of one and the same musical work or to two performances of two distinct musical works?”. They had to provide their response on a scale from 1 to 5:

- (1) Definitely the same work;
- (2) Probably the same work;
- (3) I don’t know/I can’t tell;
- (4) Probably two distinct works;
- (5) Definitely two distinct works.

2.1. Study 1a. Ukraine and Portugal

The story in Study 1a contains a change in descriptive purpose and is adapted from Puy (2019, 243):

[*Intro*] Imagine a composer who wrote a symphony with the purpose of recreating Ukrainian atmosphere. With that goal in mind, he used many Ukrainian folk melodies in the symphony. A second composer decided to rewrite the symphony

[*Purpose retained*] maintaining the same purpose to depict Ukrainian atmosphere, but he replaced some Ukrainian folk tunes with other traditional Ukrainian melodies.

[*Purpose changed*] and to change its purpose. Instead of depicting Ukrainian atmosphere, the second composer decided to recreate Portuguese atmosphere. All the Ukrainian folk tunes were replaced with Portuguese folk tunes.

2.2. Study 1b. Formal purpose

Materials. The story in Study 1b describes a change in formal purpose (developing experimental harmonies, as proposed by Puy (2019, 244)):

[*Intro*] Imagine a composer who wrote a symphony with the purpose of developing new experimental harmonies. Since the symphony was extremely hard to perform, a second composer decided to rewrite it, making some parts easier to play. The second composer [*Purpose retained*] kept the same purpose of the symphony – listeners are still surprised by new and unheard experimental harmonies.

[*Purpose changed*] didn't want harmonies to sound new and unheard. Instead of that purpose, his music embodied a new one: to imitate harmonic structures of music written by Debussy almost one hundred years ago.

2.3. Study 1c. Beethoven

Materials. The purpose in this story is again a descriptive one, but the described object is a character.

[*Purpose retained*] Beethoven composed his third symphony with a purpose to “celebrate the memory of a great man”. It was published in 1806. However, the symphony had been performed before publishing.

[*Purpose changed*] In 1803-1804, Beethoven composed his third symphony with a purpose to embody Napoleon's heroism and the ideals of the French revolution (liberty, equality

and brotherhood). He entitled it “Bonaparte”. A year later, after hearing that Napoleon had declared himself Emperor of France, Beethoven tore up the title page, eliminated the title “Bonaparte” and gave the piece a new title, “Heroic Symphony”. The purpose of the symphony was no longer to celebrate Napoleon or the ideals of the French Revolution, but changed to “celebrate the memory of a great man” without mentioning any specific person.

2.4. Study 1d. Funeral march

Materials. In this case, the purpose is to realize a musical genre.

[*Intro*] Imagine a composer who wrote a musical work. It was a funeral march, written with the purpose to honor the composer’s father after his death. The piece was performed only once and later it was lost and forgotten. Fifty years later, another composer found the lost

[*Purpose retained*] score and tried to identify what was the original purpose of the work. He correctly identified and widely announced that it was a funeral march written with the purpose to honor the first composer’s father after his death. The second composer released the score again.

[*Purpose changed*] score. He didn’t know the original purpose of the musical work, but it sounded very melancholic and the second composer decided and widely announced that the purpose of the work must be to depict the melancholic winter landscape. The second composer released the score again.

2.5. Study 1e. Winter landscape and space exploration

Materials. In this study, the stories contain a change in the descriptive purpose with a controlled amount of changes to the score.

[*Intro*] Imagine a composer who wrote a musical work. It was written with the purpose to depict a melancholic winter landscape. Both the composer and the critics agree that the work does a good job in doing exactly this – depicting a melancholic winter landscape. Ten years later, the composer decides to try some modifications to this material in order to see whether this would

[*Purpose retained*] still allow to achieve the original purpose - to depict a melancholic winter. He ends up changing approximately one tenth of the score. Looking at the results, both the composer and the critics agree that the resulting work still does a good job in depicting a melancholic winter landscape.

[*Purpose changed*] allow to achieve a different purpose - to depict a scene of space exploration. He ends up changing approximately one tenth of the score. Looking at the results, both the composer and the critics agree that the resulting work no longer does a good job in depicting a melancholic winter landscape, but that now it does a good job in depicting a scene of space exploration.

2.6. Study 1f. Piano trio “Death”

Materials. In this study, we aimed to capture a minimal change in purpose, so we created a pair of vignettes about a musical work telling about the same process – death – and differing only in the character who undergoes that process.

[*Intro*] There was a composer who created a piano trio (musical work written to be performed on a piano, violin and cello) called “Death”. The piano trio “Death” was written with the purpose of telling the story of a character created by the Greek philosopher Hegesias. It was the story of a man who resolved to starve himself to death. The composer believes that the work perfectly fulfils this purpose and that when listeners hear “Death” they really experience the story of the man who starved himself to death. Thirty years later, the composer

[*Purpose retained*] considers whether he should modify the work. However, he still thinks that the work perfectly fulfils its purpose of telling the story of the man who starved himself to death. Thus, the composer doesn’t change the score.

[*Purpose changed*] decides that the purpose of the work is not really to tell the story of the man who starved himself to death. Instead, he decides and announces that the real purpose of the piano trio “Death” is to tell the story of Giordano Bruno who was burned to death. However, the composer doesn’t change the score.

Results. For purposes of statistical analysis, the data on individuation in studies 1a-f were treated as ordinal data and “I don’t know/I can’t tell” was treated as the middle-point of this scale. Mann-Whitney U tests were conducted for pairwise comparisons. Rank biserial correlations are

reported as effect size measures. Distribution of responses and results of statistical tests are presented in Table 2. Results are presented in Figure 1.

		Definitely the same musical work	Probably the same musical work	I don't know/I can't tell	Probably two distinct works	Definitely two distinct works	Difference from midpoint
Study 1a	Retained	3.9%	7.9%	13.2%	50.0%	25.0%	$p<.001$
	Changed	1.4%	4.1%	5.5%	43.8%	45.2%	$p<.001$
$U=2062, p=.003, r_{rb}=.26$; as hypothesized							
Study 1b	Retained	7.5%	31.3%	4.5%	43.3%	13.4%	$p=.123$
	Changed	0%	22.4%	9.0%	47.8%	20.9%	$p<.001$
$U=1825, p=.047, r_{rb}=.19$; as hypothesized							
Study 1c	Retained	23.6%	43.4%	13.2%	17.9%	1.9%	$p<.001$
	Changed	46.2%	35.0%	8.5%	8.5%	1.7%	$p<.001$
$U=4578, p<.001, r_{rb}=-.26$; contrary to the hypothesis							
Study 1d	Retained	24.0%	39.4%	9.6%	19.2%	7.7%	$p<.001$
	Changed	29.3%	31.0%	8.6%	23.3%	7.8%	$p<.001$
$U=5995, p=.936, r_{rb}=.01$; no difference between conditions							
Study 1e	Retained	11.9%	31.2%	3.7%	35.8%	17.4%	$p=.227$
	Changed	6.5%	14.1%	7.6%	56.5%	15.2%	$p<.001$
$U=4197, p=.035, r_{rb}=.16$; as hypothesized							
Study 1f	Retained	68.0%	24.0%	4.0%	2.0%	2.0%	$p<.001$
	Changed	52.3%	24.2%	8.6%	13.3%	1.6%	$p<.001$
$U=4161, p=.004, r_{rb}=.19$; as hypothesized							

Table 2. Distribution of responses in studies 1a-f. Median response option is enhanced in bold. "Difference from midpoint" column indicates whether the Wilcoxon signed rank tests indicate that results are statistically significantly different from the midpoint ("I don't know/I can't tell"). For each study, results of Mann-Whitney U test of are presented.

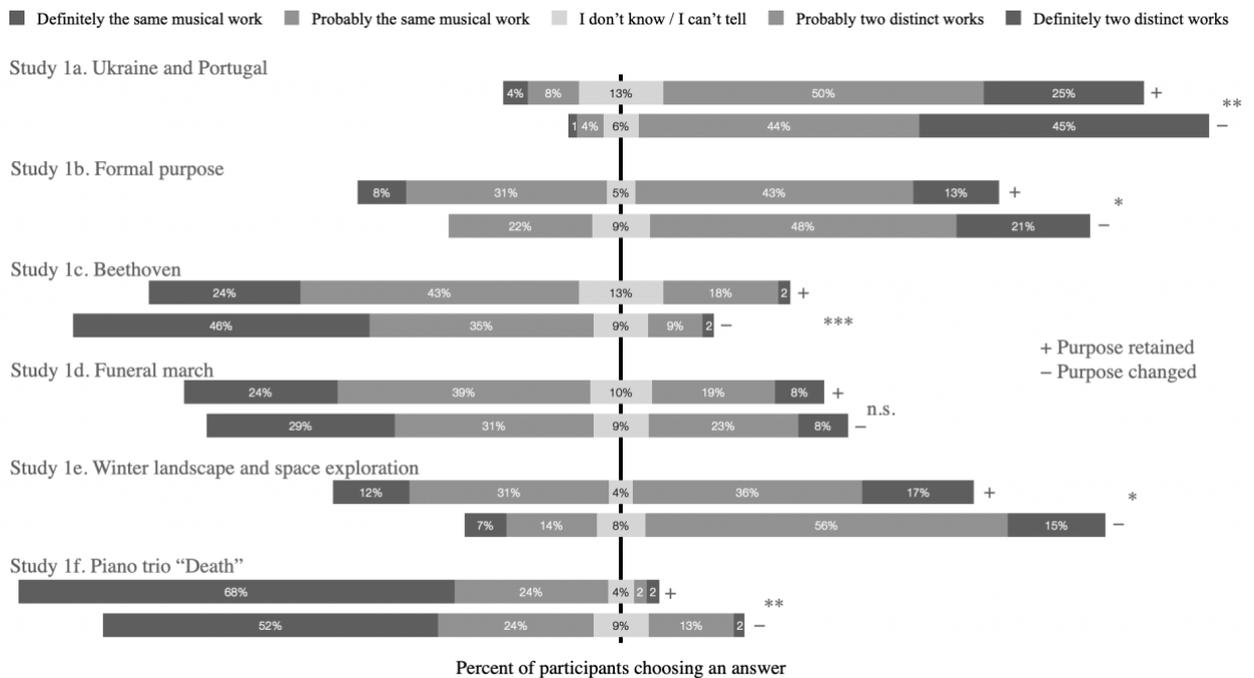


Figure 1. Percent of participants in studies 1a-f choosing an answer. Divided by condition: purpose retained (+) and purpose changed (-). Significance levels: n.s. $p > .05$; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

As summarized in Table 2, in four out of six studies (1a, 1b, 1e, 1f), purpose had the hypothesized effect: participants were more inclined to think there are two distinct works when the purpose was changed. In all four cases, however, the strength of association was weak, $r_{rb} = .16$ -. $.26$. Purpose had no effect in 1d and even had a weak effect ($r_{rb} = -.26$) in the opposite direction in 1c.³

While participants' individuation judgments differed a lot across vignettes (from vast majority leaning towards "two distinct works" in 1a to vast majority leaning to "the same work" in 1f), in none of the studies manipulation of purpose led to a change in median response.

In addition to these analyses, we also analyzed the data of all six individual studies in combination using an ordinal logistic regression.⁴ A cumulative link model was estimated (using R package "ordinal") to investigate whether the changes in purpose, author, and score predict the judgments about individuation of musical works. Wald tests showed that each of the three predictors was significant while controlling for the effect of the remaining two: Purpose, $b = .223$, $SE = .109$, $z = 2.047$, $OR = 1.25$, 95% CI [1.01, 1.55], $p = .041$, Author, $b = .795$, $SE = .114$, $z = 6.958$, $OR = 2.21$, 95% CI [1.77, 2.77], $p < .001$, and Score, $b = 2.059$, $SE = .123$, $z = 16.312$, $OR = 7.84$, 95% CI [6.12, 10.03], $p < .001$. When score is changed, the odds of the individuation rating being closer to the "two works" end of scale is 7.84 times that of when score is preserved, an effect of a large size (Chen, Cohen and Chen 2010). Odds ratio was small for the change in author (2.21) and very small for the change in purpose (1.25).

Discussion. In four out of six cases (1a, 1b, 1e, 1f), purpose had the hypothesized effect: participants were more inclined to think there are two distinct works when the purpose was changed. However, purpose had no effect in 1d and even had an opposite effect in 1c. The reason for purpose having no effect in 1d might be that changes of purpose are not considered to be real changes of musical works if they happen without changes in score. We may speculate that the opposite effect in 1c could have been caused by the historicity of the Beethoven story and it being well known that despite changes, this work is still considered as Beethoven's Third Symphony.

In studies where the score was changed, participants were generally leaning to treat the performances as of two distinct works while in studies that did not contain changes in the score, the participants were treating the two performances as performances of one work. These results draw our attention to the importance of the score (and, consequently, acoustical properties) in the individuation of musical works. Furthermore, collapsing across all six studies reported in this part, ordinal logistic regression model suggests that the change in score has a large effect on individuation judgments while the effect of author and purpose is, while detectable, much less notable.

3. Study 2

Study 1 was performed as an online study with English-speaking participants. In Study 2, we aim to explore the effect of teleology in a different population: Lithuanian-speakers. In this study, we also included professional musicians, to study whether there is any impact of musical education on intuitions about individuation.

Participants. 341 Lithuanian-speaking participants were recruited at Vilnius University ($n=255$) and Lithuanian National Opera and Ballet Theatre ($n=86$). 61% identified as females, 38% identified as males, 1% identified as non-binary. $M_{age}=25.2$; age $SD=11.1$; age range 18-62. 49% of participants indicated that they did not have musical education, 11% indicated that they have less than 7 years of musical education, 39% indicated that they have 7 or more years of musical education, 1% did not provide their response concerning musical education.

Methods and materials. Participants were divided into two groups. One group received three vignettes (those from studies 1a, 1c, and 1e) with purpose retained, another – the same three vignettes with purpose changed. Vignettes were presented in two counterbalanced orders: *ace* and *eca*. Participants completed the questionnaire using pen and paper in auditorium.

Results. No order of presentation effects were observed, so the data were collapsed across orders for further analysis. Distribution of responses and results are presented in Table 3 and Figure 2.

		Definitely the same musical work	Probably the same musical work	I don't know/I can't tell	Probably two distinct works	Definitely two distinct works	Difference from midpoint
Ukraine	Retained	3.4%	10.3%	1.1%	51.7%	33.3%	$p < .001$
	Changed	1.2%	11.4%	2.4%	47.9%	37.1%	$p < .001$
$U=13997, p=.522, r_{rb}=.04$; no difference between conditions							
Beethoven	Retained	42.5%	35.1%	2.3%	18.4%	1.7%	$p < .001$
	Changed	32.9%	42.5%	3.0%	18.0%	3.6%	$p < .001$
$U=13234, p=.130, r_{rb}=.09$; no difference between conditions							
Winter	Retained	4.0%	24.1%	5.7%	43.1%	23.0%	$p < .001$
	Changed	2.4%	11.4%	3.0%	42.5%	40.7%	$p < .001$
$U=10919, p < .001, r_{rb} = .25$; as hypothesized							

Table 3. Distribution of responses in Study 2. Median response option is enhanced in bold. Difference from midpoint column indicates whether the Wilcoxon signed rank tests indicate that results are statistically significantly different from the midpoint (I don't know/I can't tell). For each vignette, results of Mann-Whitney U test of are presented.

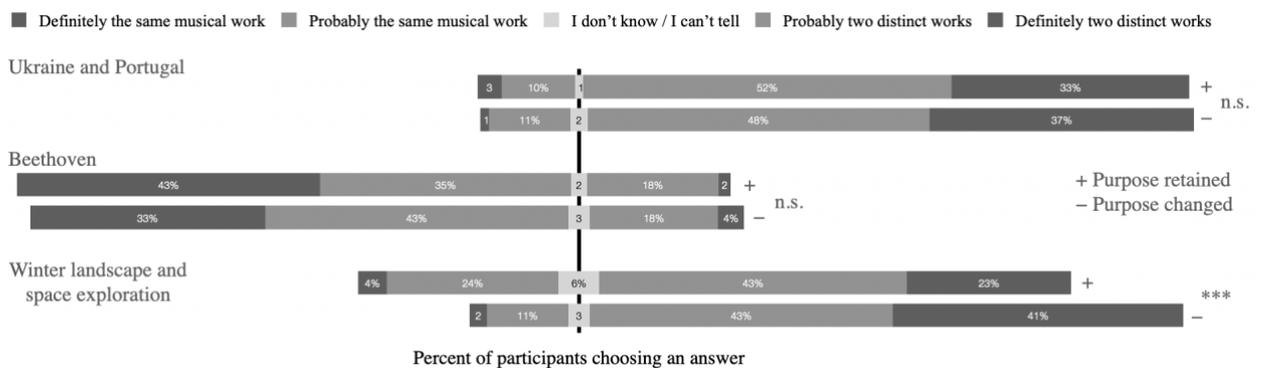


Figure 2. Percent of participants in Study 2 choosing an answer. Divided by condition: purpose retained (+) and purpose changed (-). Significance levels: n.s. $p > .05$; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

As summarized in Table 3, purpose had the hypothesized effect only in “Winter landscape and space exploration”: participants were more inclined to think there are two distinct works when

purpose was changed. This effect, however, was weak ($r_{rb}=.25$) and it did not result in change in median answer. Purpose had no effect in other two vignettes.

No statistically significant differences in individuation judgments for any of the six versions of the vignettes were observed between those who indicated 7 or more years of musical education and those who indicated no musical education. The general pattern of responses (statistically significant difference between conditions in “Winter landscape and space exploration” (musicians: $p=.008$, $r_{rb}=.25$; no musical education: $p=.002$, $r_{rb}=.26$), no difference in other two vignettes) holds for both groups taken separately.

We used cumulative link mixed model with participants as a random factor to investigate whether changes in purpose, author, and score predict judgments about individuation of musical works. Wald tests suggest that each of the three predictors was significant while controlling for the effect of the remaining two: Purpose, $b=.467$, $SE=.135$, $z=3.571$, $OR=1.60$, 95% CI [1.24, 2.07], $p<.001$, Author, $b=.350$, $SE=.146$, $z=2.406$, $OR=1.42$, 95% CI [1.07, 1.89], $p=.016$, and Score, $b=2.793$, $SE=.177$, $z=15.813$, $OR=16.32$, 95% CI [11.55, 23.07], $p<.001$. When score is changed, the odds of the individuation rating being closer to the “two works” end of scale is 16.32 times that of when score is preserved, an effect of a large size. Odd ratios were very small for both change in purpose (1.60) and author (1.42).

Discussion. The purpose had hypothesized effect in only one out of three cases. This suggests that the effect of purpose is not strong or reliable. Once again, collapsing across the three pairs of vignettes reported in this part, the change in score had a large effect on individuation judgments while the effects of, respectively, author and purpose, while present, were much less notable.

Study 3

In this study, we attempt to look at three factors that potentially influence judgments of individuation in a systematic manner: change in purpose, score, and composer. Our intention is to provide more context for evaluation of strength of impact of teleology on individuation.

Participants. 139 Lithuanian-speaking participants were recruited at Vilnius University ($n=68$) and Lithuanian Academy of Music and Theatre ($n=71$). 57% identified as females, 42% identified as males, 1% identified as non-binary. $M_{age}=20.4$; age $SD=1.40$; age range 18-27. 35% of participants indicated that they did not have musical education, 17% indicated that they have less than 7 years of musical education, 48% indicated that they have 7 or more years of musical education. 1 participant did not provide demographic information.

Methods and materials. Each participant received six vignettes in Lithuanian language. All six vignettes were versions of “Winter landscape and space exploration” (1e) and they were identical except the changes to the three variables: (a) purpose either changed or retained; (b) score either changed or retained; (c) change made by either the same or a different composer. This would result in 8 possible combinations, but we dropped two (in which both purpose and score is retained), because it does not make sense to ask about potential change of work’s identity when no changes were introduced. Participants completed the questionnaire using pen and paper in auditorium.

Here is one example of the vignette used in this study (same composer, changes to purpose and score):

Imagine a composer who wrote a musical work. It was written with the purpose to depict a melancholic winter landscape. Both the composer and the critics agree that the work does a good job in doing exactly this – depicting a melancholic winter landscape. Ten years later, the composer decides to try some modifications to this material in order to see whether this would allow to achieve a different purpose – to depict a scene of space exploration. He ends up changing approximately one tenth of the score. Looking at the results, both the composer and the critics agree that the resulting work no longer does a good job in depicting a melancholic winter landscape but that now it does a good job in depicting a scene of space exploration.

Results. Distribution of responses is provided in Table 4.

Vignette	Purpose	Score	Author	Definitely the same musical work	Probably the same musical work	I don't know/I can't tell	Probably two distinct works	Definitely two distinct works	Difference from midpoint
1	Same	Different	Different	4.3%	30.9%	6.5%	35.3%	23.0%	$p<.001$
2	Different	Different	Different	3.6%	10.1%	5.8%	34.5%	46.0%	$p<.001$
3	Same	Different	Same	12.2%	40.3%	7.9%	28.1%	11.5%	$p=.244$
4	Different	Different	Same	5.8%	13.7%	7.2%	45.3%	28.1%	$p<.001$
-	Same	Same	Different	-	-	-	-	-	-
5	Different	Same	Different	43.9%	25.9%	8.6%	14.4%	7.2%	$p<.001$
-	Same	Same	Same	-	-	-	-	-	-
6	Different	Same	Same	46.8%	25.2%	8.6%	15.1%	4.3%	$p<.001$

Table 4. Distribution of responses in Study 3. Median response option is enhanced in bold. Difference from midpoint column indicates whether the Wilcoxon signed rank tests indicate that results are statistically significantly different from the midpoint (I don't know/I can't tell).

Effect	Comparison	<i>Mdn_{same}</i>	<i>Mdn_{different}</i>	<i>Z, p</i>	<i>r_{rb}</i>
Purpose	Different score, different author (1vs2)	4	4	-5.75, <i>p</i> <.001	.31
	Different score, same author (3vs4)	2	4	-6.37, <i>p</i> <.001	.38
Score	Different purpose, different author (5vs2)	2	4	-8.34, <i>p</i> <.001	.70
	Different purpose, same author (6vs4)	2	4	-8.37, <i>p</i> <.001	.65
Author	Same purpose, different score (3vs1)	2	4	-4.55, <i>p</i> <.001	.24
	Different purpose, different score (4vs2)	4	4	-2.79, <i>p</i> =.004	.19
	Different purpose, same score (6vs5)	2	2	-.72, <i>p</i> =.466	.04

Table 5. Pairwise comparisons for key vignette combinations in Study 3. For each comparison, results of Mann-Whitney U test of are presented.

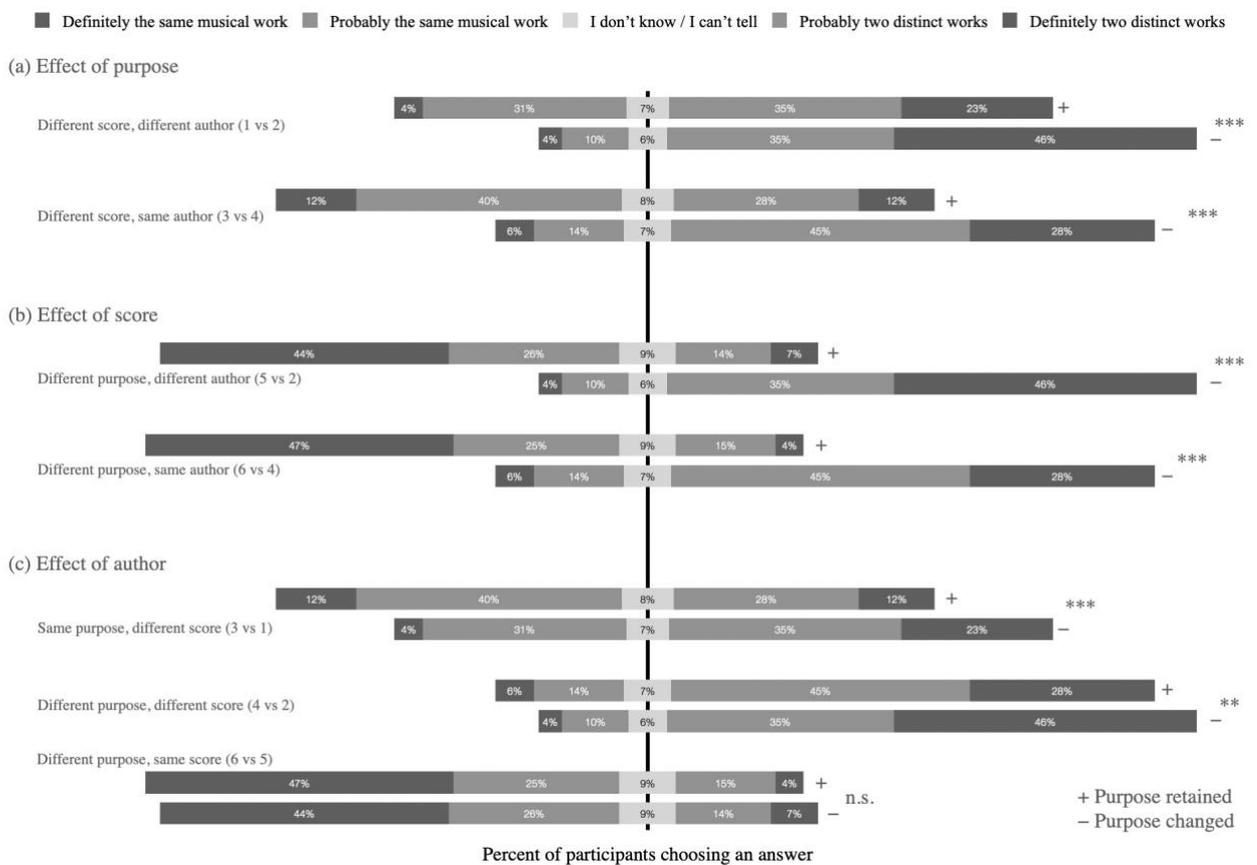


Figure 3. Percent of participants in Study 3 choosing an answer. Divided by condition: relevant characteristic (purpose, score, author) retained (+) and changed (-). Note that the same condition may appear in different comparisons. Significance levels: n.s. *p*>.05; * *p*≤.05; ** *p*≤.01; *** *p*≤.001.

As in Study 2, cumulative link mixed model with participants as a random factor was used to investigate whether changes in purpose, author, and score predict judgments about individuation of musical works. Wald tests suggest that each of the three predictors was significant while controlling for the effect of the other two: Purpose, *b*=1.300, *SE*=.164, *z*=7.914, *OR*=3.67, 95% CI

[2.66, 5.07], $p < .001$, Author, $b = .590$, $SE = .133$, $z = 4.434$, $OR = 1.80$, 95% CI [1.39, 2.34], $p < .001$, and Score, $b = 3.175$, $SE = .199$, $z = 15.974$, $OR = 23.93$, 95% CI [16.21, 35.33], $p < .001$. When score is changed, the odds of the individuation rating being closer to the “two works” end of scale is 23.93 times that of when the score is preserved, a large effect. Odd ratio was of medium size for the change in purpose (3.67) and small for the change in author (1.80).

In order to further explore the impact of our manipulations of characteristics (purpose, score, author) on judgments of individuation, we conducted a series of pairwise comparisons, looking into pairs of vignettes that differ in only one parameter. See Table 5 for the summary results. In order to get a strength of association measure comparable to Studies 1 and 2, we calculated rank biserial correlations between individuation judgments and condition (i.e., whether the relevant parameter was retained or changed).

Purpose. Two pairwise comparisons address the issue whether the changing vs retaining purpose will have an influence on individuation (Figure 3a). In both available pairs that differed only in terms of purpose, manipulation of purpose had an effect: participants were more leaning to judge that the story describes one rather than two works when the purpose was retained (both $ps < .001$). Strength of association in both cases was moderate ($r_{rb} = .31$ and $.38$). In one of these pairs (different score, same author), manipulation of purpose resulted in change in median response. Median response when the purpose was retained was 2 (there was no statistically significant difference from the middle of the scale, $p = .244$, however), while when the purpose was changed, the median response was 4, significantly leaning towards treating the two performances as performances of two distinct works ($p < .001$).

Score. In both available pairs that differed only in terms of score (Figure 3b), manipulation of score had an effect: participants were more leaning to judge that the story describes one rather than two works when score was not changed (both $ps < .001$). Association in both cases was strong ($r_{rb} = .70$ and $.65$). In both pairs, manipulation of score resulted in change in median response. Median response when score was retained was 2, while when the score was changed, the median response was 4. In all cases there was a statistically significant difference from the middle of the scale (all $ps < .001$).

Author. Statistically significant effect was observed in two cases out of three ($ps < .005$), such that change by the same author is perceived as more identity-preserving than a change introduced by a different author (Figure 3c). Strength of association in both cases was, however, weak ($r_{rb} = .24$ and $.19$). In none of the pairs manipulation of author resulted in change of median response.

No statistically significant differences in individuation judgments for five of the six vignettes were observed between those who indicated no musical education and those who

indicated 7 or more years of musical education. However, there was a significant difference for one vignette: different author, different purpose, same score. For this vignette, Mann-Whitney U test showed that participants with no musical education were more leaning to judge that the story describes one rather than two works than participants who indicated 7 or more years of musical education ($Mdn_{non-musicians}=1$, $Mdn_{musicians}=2$, $n=112$, $U=1119$, $p=.005$, $r_{rb}=.29$). However, both groups treated this vignette as describing two performances of the same musical work (both $ps<.001$).

Discussion. In study 3, the effect of purpose was observed, although it did not change the results as much as the changes in score did. The same pattern of responses was observed as in the previous two studies: the change in score had a large effect on individuation judgments while the effects of, respectively, author and purpose, while present, were much less notable.

4. Conclusion

We conducted three studies to examine the role of teleological thinking in judgments on the persistence of musical works. Overall, the results support weak teleological hypothesis. However, the effect of teleology is relatively small and was observed only in some scenarios. A strong version of the teleological view, or Puy's hypothesis, which claims that different versions must keep the overall point or purpose of the work to be considered as versions of the same musical work, is not supported by our results.

Puy discusses the distinction between "different versions of the same work" and "different works". We did not explicitly ask participants about versions and presented a more general distinction of "the same work" and "two distinct works". Although it is hard to say whether results would have been different if participants were explicitly asked about versions, our results are nonetheless directly relevant to Puy's hypothesis, as it predicts that continuity of the purpose is necessary for work's identity. Our results contradict Puy's hypothesis. However, they support a weak version of teleological view, which claims that study participants will be more inclined to believe that there is one and the same musical work if the purpose is retained.

In addition, it seems that changes of purpose are admitted as changes to identity only if they happen together with changes in the score. In all three studies, the same pattern of responses was observed: the change in score had a large effect on individuation judgments while the effects of, respectively, author and purpose, while present, were much less notable. Talking about the effect of purpose in particular, a weak to moderate effect of purpose was found in 4 cases out of 6

in Study 1, in 1 case out of 3 in Study 2, and in 2 cases out of 2 in Study 3. No notable differences were found between ratings of musicians and non-musicians.

While our main goal was to examine the significance of purpose for judgments of persistence of musical works, there are additional interesting findings on the effects of score and composer. First, the results show consistent support for the effect of score on the persistence of musical works. The effect of score is considerably larger than of the other two factors and supports a strong version of sonicist hypothesis. The importance of score in judgments on the identity of musical works echoes the results of earlier studies on individuation (Mikalonytė and Dranseika 2020) and gives some support to intuitiveness of sonicism (Dodd 2007) in the discussion on the individuation of musical works. Moreover, intuition that changes in score result in a distinct work gives some support to Dodd's (2007) view that musical works are temporally inflexible.

If we interpret the results in the light of disagreement between Puy's and Dodd's positions on works' identity, our results seem to be considerably more compatible with the latter. One limitation of the present study, however, is the use of vignettes instead of acoustic stimuli. Because of the lack of acoustic information, the significance of perceptual factors might have been underestimated, while the significance of conceptual (and teleological) factors might have been overestimated. Although the relative importance of perceptual and conceptual factors cannot be reliably compared, having in mind earlier empirical studies on perceptual and conceptual properties of visual artworks, our initial findings suggest that perceptible properties of music may be more important in individuation process than in the case of visual art. In the case of aural artworks, aesthetic empiricism (Levinson 2017, 20) might possibly turn out to be a sounder theory than aesthetic contextualism (Dodd 2007, 6). We hope that future research will investigate these potential differences more extensively.

Finally, the effect of composer appears to be very weak. Our results in Study 3 offer some modest support for a weak version of the hypothesis that change by a second composer is more disruptive to work identity than change by the original composer (Friedell 2020). However, our main goal was to see whether changing the purpose has an influence on judgments of persistence, and we gave less attention to the other two variables. Thus, more research is needed to explore these findings further.

The significance of our main results, showing the moderate effect of teleology, will depend on our perspective. Looking from the viewpoint of ontologists of music, they may want to take into consideration that folk intuitions on musical works are somewhat susceptible to teleological tendencies. However, for those working in experimental philosophy of metaphysics or psychology of art, this effect may look smaller than one could have expected.

In ontology of musical works, our results may lead to two different kinds of implications. Those ontologists who aim to establish a dialogue between ontology of art and general contemporary metaphysics and who seek to find ontological categories common to objects of art and objects from other domains, should take into consideration the teleological element in folk intuitions and seek to eliminate it from the identity conditions of musical works. This position would lead to revision of our folk theory. The authors of this study lean towards this view, although appropriate argumentation for it is beyond the scope of this article. We think that the teleological element of our folk theories should be eliminated – regardless of whether we are talking about artifacts or natural objects.

By contrast, those ontologists who are content with working with categories confined to the field of art, especially those taking the methodological stance of formulating descriptive theories, should incorporate the teleological element into their theories in a similar way as was proposed by Puy. However, they should consider the fact that teleology does not seem to play a very important role in our folk-musicological theory.⁵

Bibliography

1. Bloom, Paul. 1996. "Intention, History, and Artifact Concepts." *Cognition* 60: 1–29.
2. Chen, Henian, Patricia Cohen, and Sophie Chen. 2010. "How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies." *Communications in Statistics—Simulation and Computation* 39(4): 860–864.
3. Dink, Jacob Q., and Lance J. Rips. 2017. "Folk Teleology and its Implications." In: *Experimental Metaphysics*, edited by David Rose, 207–236. London: Bloomsbury.
4. DiYanni, Cara, and Deborah Kelemen. 2005. "Time to Get a New Mountain? The Role of Function in Children's Conceptions of Natural Kinds." *Cognition* 97(3): 327–335.
5. Dodd, Julian. 2007. *Works of Music: An Essay in Ontology*. Oxford: Oxford University Press.
6. Dutton, Denis. 2009. *The Art Instinct: Beauty, Pleasure, & Human Evolution*. New York: Oxford University Press.
7. Friedell, David. 2020. "Why Can't I Change Bruckner's Eighth Symphony?" *Philosophical Studies* 177: 805–824.
8. Gelman, Susan A. 2013. "Artifacts and Essentialism." *Review of Philosophy and Psychology* 4: 449–463.

9. Goodman, Nelson. 1968. *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis and New York: The Bobbs-Merrill Company.
10. Hume, D., 2007. *A Treatise of Human Nature: Volume I*. Edited by: David Fate Norton, Mary J. Norton. Oxford: Clarendon Press.
11. Kelemen, Deborah. 1999. "Why Are Rocks Pointy? Children's Preference for Teleological Explanations of the Natural World." *Developmental Psychology* 35(6): 1140–1452.
12. Kelemen, Deborah, Joshua Rottman, and Rebecca Seston. 2013. "Professional Physical Scientists Display Tenacious Teleological Tendencies: Purpose-Based Reasoning as a Cognitive Default." *Journal of Experimental Psychology: General* 142(4): 1074–1083.
13. Kelemen, Deborah, and Evelyn Rosset. 2009. "The Human Function Compunction: Teleological Explanation in Adults." *Cognition* 111(1): 138–143.
14. Levinson, Jerrold. 2017. *Aesthetic Pursuits: Essays in Philosophy of Art*. Oxford: Oxford University Press.
15. Mag Uidhir, C. 2013. "Art, Metaphysics, and the Paradox of Standards." In: *Art and Abstract Objects*, edited by Christy M. Uidhir, 1–26. Oxford: Oxford University Press.
16. Medin, Douglas L., and Andrew Ortony. 1989. "Psychological Essentialism." In: *Similarity and Analogical Learning*, edited by Styliani Vosniadou and Andrew Ortony, 179-195. New York: Cambridge University Press.
17. Mikalonytė, Elzė S., and Vilius Dranseika. 2020. "Intuitions on the Individuation of Musical Works. An Empirical Study." *British Journal of Aesthetics* 60(3): 253–282.
18. Newman, George E. 2016. "An Essentialist Account of Authenticity." *Journal of Cognition and Culture* 16: 294–321.
19. Newman, George E., Daniel M. Bartels, and Rosanna K. Smith. 2014. "Are Artworks More Like People Than Artifacts? Individual Concepts and Their Extensions." *Topics in Cognitive Science* 6: 647–662.
20. Newman, George E., and Paul Bloom. 2012. "Art and Authenticity: The Importance of Originals in Judgments of Value." *Journal of Experimental Philosophy: General* 141(3): 558–569.
21. Newman, George E., and Rosanna K. Smith. 2018. "Artworks Are Evaluated as Extensions of Their Creators." In: *Advances in Experimental Philosophy of Aesthetics*, edited by Florian Cova and Sebastien Rehault, 103–120. London and New York: Bloomsbury.
22. Puy, Nemesio G. C. 2019. "The Ontology of Musical Versions: Introducing the Hypothesis of Nested Types." *The Journal of Aesthetics and Art Criticism* 77(3): 241–254.

23. Rabb, Nathaniel, Hiram Brownell, and Ellen Winner. 2018. "Essentialist Beliefs in Aesthetic Judgments of Duplicate Artworks." *Psychology of Aesthetics, Creativity, and the Arts* 12(3): 284–293.
24. Rose, David. 2015. "Persistence Through Function Preservation." *Synthese* 192: 97–146.
25. Rose, David, and Shaun Nichols. 2019. "Teleological Essentialism." *Cognitive Science: A Multidisciplinary Journal* 43: 1–19.
26. Rose, David, and Shaun Nichols. 2020. Teleological Essentialism: Generalized. *Cognitive Science: A Multidisciplinary Journal* 44: e12818.
27. Rose, David, and Jonathan Schaffer. 2017. "Folk Mereology is Teleological." *Nous* 51(2): 238–270.
28. Stavrova, Olga, George E. Newman, Anna Kulemann, and Detlef Fetchenhauer. 2016. "Contamination Without Contact: An Examination of Intention-Based Contagion." *Judgment and Decision Making* 11(6): 554–571.

¹ For more on the distinction between these two positions (called, respectively, Deference view and Independence view), see Uidhir (2013).

² So, while weak hypotheses claim that the variable of interest will have a statistically significant effect on individuation judgments, strong hypotheses include an additional claim that the change in the variable of interest will result in the reversal of judgments of individuation (e.g. from overall agreement with the claim that there are two musical works to overall agreement with the claim there is one musical work).

³ In order to check whether the purpose manipulation works, in Study 1f an additional question was asked: "Please indicate to what extent you agree or disagree with the following claim: "The musical work "Death" when it was originally written had the same purpose as the musical work "Death" thirty years later" (answered on 7-point Likert scale anchored at (1) Completely disagree and (7) Completely agree). Results suggest that the manipulation was highly effective: $M_{retained}=6.39$, $SD=.97$, $M_{changed}=3.30$, $SD=2.09$, $t(226)=13.7$, $p<.001$, $d=1.82$.

⁴ We would like to thank an anonymous reviewer for this suggestion.

⁵ We thank Evelina Greiciūnaitė for her assistance with data collection.