

University of Groningen

Beauty in mathematics

Montano Juarez, Ulianov

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Montano Juarez, U. (2010). *Beauty in mathematics*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CONCLUSION

Many mathematicians have given lyric expression to their enthusiasm for mathematical beauty but very few have addressed an actual aesthetics of mathematics. In attempting this here I have tried to stretch the boundaries of aesthetics, but I can only claim to have roughly sketched the very first draft of such an aesthetics.

Following Francois Le Lionnais' call for a serious aesthetic of mathematics and Gian-Carlo Rota's concern with the use of the *term* 'mathematical beauty', I have approached beauty in mathematics by addressing mathematical aesthetic judgements. My strategy was to present an aesthetic theory that allows for a literal interpretation of mathematical aesthetic judgements. I have proposed a formalist-oriented model of the-aesthetic-as-a-process. Aesthetic-processes are collections of interrelated events, the development of these processes involve the interaction of objects, subjective reactions, and historic-cultural contexts. The idea of the-aesthetic-as-a-process provides a general framework for understanding aesthetic events in mathematics as a particular class of aesthetic events. I have focused on the elements relevant to understanding aesthetic experience, value and judgements in mathematics. I have interpreted aesthetic experience as an embedded process which is unified by a mathematical intentional object (although there might be cases of incomplete or alienated experience). The experience-process develops by undergoing changes in the passive and active content and by eliciting affective responses. An aesthetic mathematical intentional object (which is an aesthetic *form*) unifies and causes the affective part of the experience. This object exists in a phenomenological space whose dimensions are properties relevant to our affective reactions. Knowledge plays a role analogous to sensory-perception since it enables us to 'see' the intentional object. Aesthetic value has been interpreted as a relation between sets of properties and affective reactions. Our evaluations are driven by preferences that manifest themselves in two ways: affectively, as affective evaluations, and cognitively, as rule-like aesthetic criteria. These aesthetic criteria can be used to track the evolution of value, the mechanism of this evolution is constrained aesthetic induction, which is a

generalization of McAllister's ideas to which the constraints (critical adequacy and robustness) imposed by the nature of our affective responses have been added. Finally, I have interpreted aesthetic judgements as aesthetic descriptions that express subjective states. Aesthetic descriptions include aesthetic predicates whose conditions of application (the existence of a mapping from a label schema to a pleasure-relation and a reorganization of our new referential domain in terms of the structure of the original domain associated with the schema) are strictly subjective. Aesthetic judgements perform the functions of articulation (identifying and clarifying our state) and exchange (conveying information about this state). Articulation is the elucidation of our subjective state in terms of the conditions of application of aesthetic terms. Exchange is the sharing of the information conveyed by the judgement. The need to express subjective states by linguistic means encourages not only the articulation of individual subjective states but also the articulation of the aesthetic-process itself. The awareness of our internal state further defines and identifies the events involved in the process as aesthetic events.

With this theory it is rather trivial to address mathematical beauty. 'Mathematical beauty' is an aesthetic term. It expresses certain subjective states caused in the observer by the engagement of his attention on a certain mathematical item.

Now, perhaps it is trivial to interpret mathematical aesthetic judgements in view of the theory discussed here, but what is not trivial is that the theory offers important insights. For example, an advantage of my approach is that, as we have seen in Chapter 4, we can account for mathematical elegance and ugliness in the same way as we account for beauty. These terms express differences in our subjective experiences, in the family of terms we employ and in the referential domains involved, from our point of view. Another insight – perhaps the most important one – is that the theory allows us to distinguish different types of aesthetic experience. A difference in type of experience amounts to a difference in the modality of aesthetic experience. The modality is determined by the kind of events that focus our attention (sensory, cognitive, or both), the kind of properties involved in our phenomenological space (for example, our appreciation of representational painting depends less on knowledge than our appreciation of mathematics, and the knowledge involved in different kinds of appreciation is also a

different kind of knowledge), the kind of activities involved during the process (some types of experience are more passive or more active), and how the content relates to the affective reactions (some experiences depend on autonomic reactions while some others depend on experience). Mathematical knowledge, as *background-understanding*, plays a central role in the modality of mathematical experience. I believe something very similar occurs in the cases of musical, literary or scientific knowledge; knowledge contributes to constituting our objects of attention.

Any insights I have achieved here are, of course, partially achieved thanks to the works from which I have taken ideas. In proposing this work I have not challenged any of the authors I reviewed; rather, I have borrowed many of their views. I have merely presented a model of aesthetic events that allows us to give all those ideas a relevant place in the depiction of the aesthetic in mathematics. Schaeffer and Scruton's ideas of the acousmatic space inspired my phenomenological space. Peter Kivy's musical formalism contributed to my idea of Aesthetic Mathematical Intentional Objects, as well as to the distinction between active and passive content of experience. Jenefer Robinson's theories of emotion and expression (which contrast dramatically with Kivy's views) contributed the very idea of a systemic approach, with plenty of empirical results to inspire my claims on experience and value, and with the idea of articulation in aesthetic judgements. Shaftesbury and Hutcheson's aesthetic principles find a place in my model as dimension of the phenomenological space, that is, as part of the elements that turn an experience into an *aesthetic* experience, since they are responsible for our affective responses. Gian-Carlo Rota's emphasis on knowledge and understanding finds a central place in my model, since mathematical knowledge, background understanding of theories, results, terms, etc., is what allows us to 'see' mathematical intentional objects in the first place. I have incorporated James McAllister's and Theo Kuipers' ideas on the aesthetic induction almost literally, just adding the appropriate generalizations and introducing the constraints imposed by empirical results. Isabel Hungerland and Peter Kivy's discussions on aesthetic terms shaped my ideas of the characteristic conditions of application of aesthetic terms and the role of aesthetic judgements as locally terminal. Perhaps my boldest contribution in this work has been to employ Nelson Goodman's

approach to metaphor to propose my interpretation of aesthetic judgements and my idea of aesthetic articulation.

Limitations

In common with any approach, the model presented here has limitations. Perhaps the most obvious is that I decided to gamble on the very ancient relationship between music and mathematics to look for clues about aesthetic events in mathematics. Hence, the resulting approach has a formalist flavour. Aesthetic forms, AMIOs, play the central role of unifying experience and, by causing affective reactions, turning it into an aesthetic experience. This allows us to introduce refined accounts of experience, as was shown in Chapter 4, in the cases of incomplete or alienated experience which account for ugliness. An approach with a less formalist flavour might allow more permissiveness with incomplete or alienated experiences.

There are also some shortcomings in the specific way I have addressed the theory. First, I have presented a detailed idea of a pleasure-relation. I believe the formal treatment I have proposed here helps to clarify the complexity of our possible affective responses. But endorsing such a detailed idea, of course, increases the probability of inaccuracies in the account. However, I believe that even if my treatment turns out to be inaccurate, an analogous notion should be used to show the complexities and variety in the pleasure response. A similar warning can be issued about the details of the process of articulation in aesthetic judgements. My idea of articulation depends on the details of the elements involved: a pleasure-relation, a schema and a referential domain (a domain of mathematical items, like proofs or theorems). It is possible that the specifics presented here will turn out to be inaccurate, but I believe that a process of articulation does actually occur and plays an important role in determining how aesthetic judgements are passed; the general idea (that our point of view undergoes a change, becoming clearer and more definite) is still plausible. The general message concerning these shortcomings is that if the specifics presented here are inaccurate I believe that one should not rush to discard the framework they flesh out.

Topics for Further Research

There are topics I have not been able to discuss in this work, mainly due to practical limitations. Among these there are problems that have not been addressed and new issues and questions that arise in light of the approach I have presented.

Most obvious among the unaddressed topics, of course, is the relationship between beauty and other philosophical problems of mathematics: truth, knowledge and the ontology of mathematics being the most conspicuous. My proposal is to some extent grounded on results related to the domain of psychological and neuro-physiological phenomena, which seem rather distant from the traditional approach to the philosophical problems of mathematics. The present work should be seen as devoted to the practice of mathematics. Mathematical practice has repercussions on many aspects of other problems, but I do not believe that all problems should be addressed from that perspective. Unfortunately, I believe, truth, knowledge and ontology are particularly resilient to such a treatment.

Mathematical Truth

The methods for securing the truth of a mathematical statement do not depend on empirical information; rather, they are based on logic. Truth is relevant to beauty in my theory: truth is a condition for mathematical items to be considered as objects in a phenomenological space. Despite this, the role of beauty seems to be relatively independent of mathematical truth. But there is an issue in which beauty and truth are more alike: as Rota has pointed out, aesthetic consideration sometimes encourage *further* proofs of a theorem. Mathematical beauty is a property that plays an objective role in encouraging mathematical work, just as truth does. The search for beauty, just like the search for truth, promotes mathematical work. Perhaps both searches are valuable in themselves. Truth is certainly a value in itself; it is one of the qualities that define mathematics as a scientific enterprise. But what about beauty? In Rota's words the search for beauty distinguishes mathematics among the sciences; in this sense mathematics has a particular stance as a cognitive scientific discipline. I believe that beauty is a

characteristic value of the mathematical discipline, just as truth is, but this issue deserves more space than that available here.

The general subject of truth poses other issues that we can briefly survey. For example, let us recall that I have incorporated McAllister's ideas in my approach. McAllister claims that the relation between truth and beauty must be clarified by empirical means (McAllister, 1996, 98-104). Theo Kuipers (Kuipers, 2002) further explores the formal relation between truth of empirical theories and aesthetic induction. In these approaches the relation between beauty and truth in science is contingent and should be investigated empirically. My version of aesthetic induction does not allow us to borrow this conclusion very easily. For example, my model incorporates an autonomic component, which is very likely determined by evolutionary adaptation; this introduces a further element to take into account. Furthermore, this element is particularly relevant in the role of specific properties like symmetry, simplicity or uniformity. The relation between truth and beauty thus also involves the relation between symmetry, simplicity, uniformity, etc. and truth. This problem is more complicated in my approach than in McAllister's.

Mathematical Knowledge and Ontology

One of the most important problems of mathematical knowledge is its very possibility (Balaguer, 1998; Colyvan, 2003; Wright, 1994). The problem of how mathematical knowledge is possible has to do with the fact that mathematical objects are abstract: they have no spatial or temporal location and they do not interact causally with the physical universe. If there is no way in which mathematical objects can interact with our physical universe, how can we have any knowledge of mathematical objects at all?

Now, the knowledge involved in my approach, background understanding, has to do with the mathematical items as objects of experience. These objects are mental events. They are as problematic as any mental content, but not necessarily in the same sense as causally isolated abstract objects. My discussion of beauty does not add much regarding the relation between mental objects and knowledge of causally isolated entities. However, in this respect it does not fare differently from other approaches to mathematical beauty. Rota's approach, for example, interprets mathematical beauty as a form of knowledge.

For Rota the property of being enlightening consists in enabling us to see connections and relations that help us to have a proper grasp of the significance of a mathematical item. But this is not the kind of knowledge that concerns the problem above.

Concerning the ontological issues of mathematics, they are closely linked to the problem of knowledge (Colyvan, 2003). Under a realist interpretation of mathematics, mathematical theories refer to abstract objects. I have explicitly differentiated mathematical objects from mathematical items, so it would seem that our discussion has little to contribute.

However, the problems related to the existence and properties of abstract objects differ depending on which approach we take. In a realist approach to mathematical objects, mathematical items and objects are relatively independent of each other. In a nominalist approach, like Hartry Field's (Field, 1980), a closer contact can be established between them. In Field's nominalism, mathematical objects do not literally exist and mathematical theories and statements are true merely in the same sense in which we say that a narrative *fiction* is true. Field's nominalism leaves some questions unanswered, such as what kinds of 'fictions' are mathematical objects? And how exactly do they work in practice? A literal aesthetics of mathematics can surely offer clues to study some qualities of mathematical objects seen as fictions, especially because mathematical practice is driven by aesthetic considerations just as it is driven by a search for a 'fictional' truth. According to Field, (fictional) mathematical objects are merely useful in practice. In the present work one of our tenets is that some of those objects can also be enjoyable. Since they are mere fictions, a balance between the practical and the enjoyable would seem to suit a nominalist account of mathematics quite well.

Other Topics for Future Work

My strategy of putting together particular aesthetic topics to present a general theory sets up, implicitly, the possibility of transferring the theory to other fields. This transference has yet to be sketched, but here I can offer a brief list of possible topics to be developed in the future.

Aesthetics of Science

Once we have contemplated the possibility of a concrete aesthetics of mathematics (not just a pre-theoretical approach), one of the tasks that comes immediately to mind is to devise a similar aesthetics of science. McAllister's work addresses beauty in science, but I have introduced changes and generalizations that substantially modify his approach. I have replaced empirical adequacy by critical adequacy, that is, by the property of being in agreement with the standards of taste. I have also introduced the effect of autonomic reactions. Empirical adequacy is out of my picture and autonomic reactions are part of it. Regarding empirical adequacy, I believe the problem is one of generality: empirical adequacy can be seen as a special case of critical adequacy. The interpretation of autonomic reactions in terms of empirical adequacy or other factors is not clear. I cannot envisage a simple way of connecting our autonomic reactions to empirical inputs. However, as we have seen, these autonomic reactions are plastic, they can be modified by our experiences, but more work is needed before I can present an accurate idea on this. The interpretation of the more theoretical notion of robustness is complicated, too. Critical adequacy is a generalization of McAllister's ideas, but robustness derives from empirical results and the idea of robustness from systems theory. In general, interpreting the new theoretical concepts introduced here in terms of the specifics of other fields is a task that still has to be completed.

Aesthetic experience, for example, needs a more detailed account of the modality of scientific beauty, that is, the role of sensory experience, the way pleasure is elicited, the role of active attention, etc.

An interpretation of aesthetic judgement depends to some extent on our interpretation of aesthetic experience; however, I believe that the general proposal presented here can be applied to science with minor modifications. Aesthetic terms in science can be characterized in the same manner as in mathematics (by the existence of a mapping and a reorganization). Aesthetic descriptions perform the same functions (articulation and exchange) too. But empirical sciences have some distinctive features that need to be addressed. In particular, the conditions of use of aesthetic terms ask for the existence of a mapping and reorganization that are changes in our point of view. These changes in point of view are easily accomplished in mathematics, since we never

have empirical access to the mathematical objects related to our objects of attention. But in the empirical sciences we have access to objects and facts in the world. This accessibility imposes constraints on the ways in which our point of view can be altered. While in mathematics the changes of point of view are only constrained by the adequacy of our mapping and reorganization, in science, when we judge scientific objects or facts, it is not clear whether these changes in point of view are arbitrarily possible. Thus, one needs to establish a distinction between purely cognitive objects of attention, like theories, theoretical terms and other constructs, and the actual facts. Such a distinction is a topic for future reflection.

Intellectual Beauty

The application of our ideas to science also presents us with the possibility of further extending our theory to ‘intellectual beauty’ (beauty in philosophy, computer programming, or less ‘serious’ fields like chess, for example) in general. As in the case of science, I believe the general outline of my theory is correct, but the specifics need to be addressed. Since intellectual beauty does not involve empirical constraints in general, we can avoid the problems that arise in science. This leaves us with the problem of suitably interpreting experience, value and judgement. The most important interpretation is that of experience. The modality of the experience of intellectual beauty is, of course, independent of sensory experience and more purely cognitive. In general it would involve some sort of active content and the eliciting of affective response, and it would depend on passive, active and learned responses. This picture is very similar to our account for mathematics; detailing it is, of course, a task for the future.

Nonetheless, it is interesting to imagine the results of a successful theory of intellectual beauty: the boundaries between art and sciences would become less sharply defined, since the appreciation of purely conceptual objects is not very different from the kind of appreciation that occurs in some types of literature and in conceptual art. The prospect of having a way to explain things like a beautiful game of chess, a beautiful scientific theory or a beautiful philosophical concept in the same terms as we would explain music, poetry or conceptual art, would offer a way to overcome the boundaries between these disciplines. Mathematical beauty and conceptual art could be seen as

special cases of intellectual beauty. The ancient Greek view of mathematics could be rescued. Perhaps, too, enthusiastic approaches like Le Lionnais', proposing that mathematics should be seen as an art, could be pursued. I am not prepared to go that far, but I believe the approach presented here is compatible with the claim that beauty in mathematics is valuable in itself; that it is not a secondary by-product, but a genuine goal.

To conclude, I believe it is not necessary to further stress the importance of beauty in mathematical practice. There is no shortage of texts, historical or philosophical, emphasizing the importance of mathematical beauty. In addition, it is also true that one usually approaches mathematics motivated by passionate curiosity; and that a need for harmony and beauty is what drives much of the mathematicians' work. My hope is that my work leaves the door open to rethink and legitimate that passion.

Appendix 1: Development of Formalism

1. Hanslick's Formalism

The Viennese musical critic Eduard Hanslick is responsible for the first appearance of musical formalism. His 1854 book *On the Beautiful in Music* is considered the first source of articulated formalist ideas (Alperson, 2004). In his book Hanslick reacts against the then popular Romantic (emotivist) theories. Hanslick's view can be summarized by saying that expression of emotions can be a source of aesthetic value; however, inquiring into these expressions does not lead us to a better understanding of the nature of the musically aesthetic. Hanslick has two goals. First, to debunk the emotionalist views on the meaning of music; he tries to discredit the claim that the content of music should be understood as emotional content and that is what determines music's aesthetic value. Second, he wants to present a positive aesthetics for music; he puts forward an alternative view of the proper object of aesthetic judgement and criticism. He claims that "the musically beautiful, as an autonomous species of beauty" (Alperson, 2004, p. 256) is the proper object of musical aesthetics.

Hanslick supports his rejection of the emotive content of music with several arguments. Two of them are based on attacks on what he believes are popular mistakes. He argues that when people say that music has to do with feelings, they implicitly assert one of two indefensible views of the relation between music and emotions. According to the first view, the purpose of music consists in arousing certain feelings in the listener. This view is sometimes called the *causal view*. According to the second view, the feelings referred to are not aroused by music, they are the subject matter, the meaning, of music. This view is known as the *semantic view* (Alperson, 2004, pp. 257-258).

Hanslick's arguments against the causal view derive from his own beliefs about beauty and the nature of music. Philip Alperson reconstructs these arguments as follows:

1 Beauty in music, as an objective property, has no purpose beyond itself. The contemplation of the beautiful may produce pleasant feelings but these have nothing to do with musical beauty. The causal view falsely supposes that because pleasant feelings may be excited by the beautiful, the purpose of music is to arouse these feelings.

2 Subjectively, music is directed to the active imagination as “the activity of pure contemplation” (Alperson, 2004, p. 257) of the sequence of tonal forms. The arousal of feelings is a secondary and peripheral effect. The beautiful in music pleases, but we cannot learn the true nature of music by dwelling on this pleasure.

Hanslick’s third and fourth arguments against the causal view depend on his own account of ‘feeling’:

3 The causal view confuses ‘feelings’ with ‘sensations’. Sensation (what we would call perception) is a precondition to the musically beautiful. Definite feelings (what we would call emotions), such as love, jubilation, and woe, however, involve an awareness of “our mental state with regard to its furtherance or inhibition, thus of wellbeing or distress” (Alperson, 2004, p. 258). The necessity of sensations does not entail the induction of feelings in the listener.

4 Furthermore, there is no causal connection between a piece of music and the specific feelings it might arouse. Musical works affect us differently depending on many factors; this is evidenced by the different emotional characterizations offered of the same musical works by different listeners.

It is true, Hanslick concludes, that music may arouse in us feelings and moods, as the causal theorists say, but the same may be said of medical reports or our luck in a lottery. We must not be misled into thinking that such feelings have anything to do with the proper understanding of music (Alperson, 2004, p. 258).

Music can arouse emotions and moods, but that does not characterize music, since many non-musical occurrences also arouse them. Emotional arousal has nothing to do with the proper understanding of music.

The representational view, which claims that music can represent emotions like love, courage or piety in the same manner as a painting can represent the happiness of a

person, is similarly flawed (Alperson, 2004, p. 258). Music lacks the resources to achieve such representations. Hanslick claims that emotions are distinguished from other kinds of affective arousal by a particular set of concepts or judgements. Hope, for example, involves the idea of a future happy state in comparison with the present. Music can only present musical ideas, not concrete judgements, and thus, even if music can imitate the dynamics of our emotional life it cannot present us with the content that characterizes emotions. Furthermore, people might agree about the beauty of a musical piece but they cannot agree about its supposed emotional content, and without agreement about what it represents there is no representation whatsoever.

Hanslick's ideas profoundly influenced the development of 20th century aesthetics. His remarks on the nature of emotions and their semantic or causal relations with works of art have been widely discussed. In musical aesthetics his influence is so extensive that many authors think of Hanslick as The Background View (Alperson, 2004).

For Hanslick, musical beauty derives from music's "tonally moving forms" (p. 260). He sometimes links this notion to language. This represents a tremendous insight. He recognizes that music shares many of the qualities of language, but he was also aware of the fact that music is not a language. One of Hanslick's most valuable contributions was the recognition that music, even if it is in many respects language-like, does not need a semantic component, or, at least, that aesthetic appreciation of music does not depend on its semantic component. Peter Kivy makes an interesting remark about Hanslick's achievement:

...what is left of a language if you distill off its semantic component? What is left is its *syntax*: its *grammatical* component. What is left is a collection of inscriptions with rules for their correct combination: rules for the string-together of the meaningless inscriptions into grammatical, syntactically correct chains, or, if you will, 'sentences'—except that they are 'sentences' without a meaning... It seems, then, that what Hanslick was at least beginning to see is that absolute music is, as it were, a kind of syntax without a semantics: language-like but not a language. It does not convey merely a sense of order. It *does* convey that. But it is a very special kind of order: the order of syntactical structure...What Hanslick realized was that, without having a meaning, absolute music at its best, has a 'logic'—a quality of inexorable progress and direction. (Kivy, 2002, p. 63)

This passage is revealing; it suggests that music possesses a kind of ‘logical necessity’. Of course the notion of logical necessity could be applied here only as a metaphor. In that metaphorical sense, musical *movements* (‘tonally *moving* forms’ is another metaphor) from one series of sound to another seem to reflect the ‘movement’ from premises to conclusions in a logical argument. In a logical argument the fulfilment of conditions in the antecedent syntactical structure necessarily guarantees that the following structure will possess a determined quality. In music, at least in Western music, as in logic, we know the set of rules governing syntactical well-formedness; those rules would correspond to the part of music theory that describes musical scales, chord structuring, time divisions, etc. We also have rules for ‘syntactical consequence’; such rules would correspond to the part of music theory describing the progression of musical material from one musical ‘sentence’ to the next, that is, the description of chord progressions, cadences, modulations, etc. Finally, in music as in logical arguments, we know and ‘feel’ that, given a specific musical structure that fulfils certain characteristics prescribed by the appropriate rules, the following musical structure must have certain characteristics.

Now, these logico-linguistic ideas are only implicit in Hanslick; his followers were responsible for rendering them explicit. Given that the development of formalism goes beyond Hanslick, we need to review it, looking for further insights into the relation between form and beauty.

2. Edmund Gurney

The development of formalism is sometimes summarized following the works of its three most important authors: Hanslick’s *On the Beautiful in Music* (1854), Edmund Gurney’s *The Power of Sound* (1880), and Leonard Meyer’s *Emotion and Meaning in Music* (1956).

Following Hanslick’s views, Edmund Gurney presented a formalist aesthetics based on melody. He characterized musical structure as ‘ideal motion’. However, even if this idea is intuitively compelling to musicians, ‘ideal motion’ is a notion that remains vague through Gurney’s work. Furthermore, sometimes the ‘ideal’ in ‘ideal motion’ has

been interpreted as describing a mysterious, non-natural property. Despite the vagueness in his conceptualization, he proposed a more concrete, practical formalism based on the idea that music is reducible to melody. However, even if the concreteness of melody-reducibility allows many practical achievements, the idea gave rise to many critical voices, especially because it neglects many traditional issues in music and also neglects non-melodic music (the music of Richard Wagner, for example).

Melody is the basis of music, according to Gurney, and melodic form determines aesthetic value. Melodic form is the “flowing connectedness of melody” (Kivy, 2002, p. 64). The ‘flowing connectedness’ refers to the fact that a melody, a stream of pitched sound with the same characteristic timbre, must fulfil several constraints to be recognized as a musical structure. The constraints are given by the rules of harmony and music theory. The phenomenon of repetition, in particular, plays an important role in this flowing connectedness. The melody must unfold according to the standard rules of harmony; the tone-relations of a musical scale, for example. Furthermore, it must fulfil certain expectations about the recurrence of certain structures. The repetition of certain sound patterns (progressions of notes) ordered according to known harmonic rules gives shape to the melody and prevents it from unfolding in a random fashion. Subjectively, that is, in the experience of a normal, non-musically-trained listener, this organization of the flowing connectedness of melody seems to refer to the sense of ‘logical’ development of music that a normal listener hears or feels in it.

In the characterization above we have a more explicit idea related to the nature of form: the idea that there is a sense of ‘consequence’ in music. Once a specific melody unfolds, the listener ‘feels’ that the melody should continue in a certain way rather than another. Although a rough idea of ‘musical necessity’ is present in Hanslick, Gurney’s focus on the simpler notion of melody and his insistence on ‘connectedness’ allow a better appreciation of it.

With Gurney we realize that musical form, at least melodic form, has a sense of ‘logical necessity’. The same ‘logical necessity’ is present in many musical phenomena, including chord progressions, repetitions of motifs, and even, according to Peter Kivy, emotional resolution.

3. Leonard Meyer' Approach

During the last three decades formalism has experienced remarkable development. Peter Kivy is recognized as a leading figure in this development. Kivy finds the term 'enhanced formalism' (Kivy, 2002, Ch. 6) suitable to describe his position. Enhanced formalism integrates emotions into the theoretical framework and the concerns of formalism. The reconciliation of emotions and formalism started Leonard B. Meyer's work.

Let us survey some of Meyer's ideas. He makes informal use of information theory. He arranges musical events from the most expected to the most unexpected. The most expected musical event is the least informative one, and vice versa. Musical events are evaluated according to the degree to which they are expected or unexpected. Meyer thinks that most of our musical expectations are determined by musical culture, that is, by an unconscious acquaintance with the styles of music available in our cultural environment. In other words, by hearing music we acquire a general idea of what music should sound like. As we hear more music we develop a more refined sense of how music 'works'. In a certain sense, as we hear music, we learn about it: we unconsciously 'understand' the general mechanics of the music we have heard. This learning, this understanding of music, depends on the specific kind of music we have been exposed to. Our understanding of how music works determines our musical expectations; when we hear a new piece of music, or a specific piece of musical sound, we expect that this music will unfold in a fashion similar to the way that we have learned from previous hearings of music.

For Meyer, music and music appreciation do not depend on form alone, nor on (emotional) content alone. Meyer presents a view of music that synthesizes both the formalist and the semanticist view. What we appreciate in music, according to Meyer, is a pure sound structure. There is no *external* content attached to the music. Meyer believes that expectations of musical developments constitute the content of music; external content refers to content other than this musical-expectation content. Now, what is interesting in Meyer is the way he explains our appreciation of musical form: our joy of music comes from the correct balance between expectation fulfilment and surprise in the

listener, when listening to a piece of music. The expectations are shaped by our musical cultural environment. Our expectations are culturally determined, but they are still determined by music alone. Meyer integrates historical and cultural dimensions in formalism.

Meyer assigns a role to the view that content determines musical value (semanticism). We acquire an understanding of how music works as we listen to music; we can see this as a kind of musical knowledge. The musical expectations could be seen as conscious or semi-conscious events⁵⁵ in the listener. Meyer thinks there is another way to look at this. He thinks we do not necessarily experience the acquaintance with music as actual knowledge of a conscious event; we can also experience it as an emotional reaction to music. Most people are not musicians and know little or nothing about how music works. Nonetheless, everyone is able to recognize music and experience expectations; the difference is that in the absence of a refined musical knowledge we experience expectations as emotional reactions. When a certain chord progression, for example, fulfils our expectations of how that progression should unfold we ‘feel’ that the music went the way it has to. When the music turns in an unexpected way we have a different feeling, a feeling that can range between surprise and frustration. We ‘feel’ that the music deviates from the expected. Meyer thinks there is nothing wrong or right about experiencing fulfilments of expectations, whether as emotional or semi-conscious (cognitive) responses. He thinks both responses are part of the appreciation of music. In Meyer’s view, formalism and emotivism are not irreconcilable positions; on the contrary, they are both correct; they just focus on different aspects of the same phenomenon. Music appreciation is a matter of both formal appreciation and emotional arousal.

Musical events have varying degrees of informational content, according to Meyer; the more a musical event fulfils expectations the less informative it is. Meyer uses this fact to define the meaning of a musical event. He defines musical meaning as the degree of fulfilment of expectations depending on the previous musical events. As we have seen, this fulfilment can be experienced as an emotional response. Thus the content of music could be construed either as a formal content, in which case only conscious

⁵⁵ We would use terms like cognitive or semi-cognitive events, referring to mental events related to knowledge, but I shall retain the term “semi-conscious” while talking about Meyer.

(cognitive) formal expectations are involved, or as emotional content, involving emotional responses.

Appendix 2: The Notion of Form in Music

1. The lack of a Notion of Form

Although attractive, the picture offered by formalism has some gaps. Formalism links pure form and musical beauty, but the nature of the link remains obscure. This, I believe, can be interpreted in terms of three interrelated problems: first, the need for a clearer notion of form; the authors surveyed in this work do not address a conceptual analysis of the notion. Second, the aesthetics, the origin of the beauty of events like the sounds of musical instruments, musical scales, chords, etc., remains unaccounted for. Formalism does not offer a full explanation of the beauty of these events. Although Kivy acknowledges this problem, he does not address it. Third, the relation between the aesthetics of basic events, like the beauty of a single chord or a single note, and formal events, like the games postulated by Kivy, is not addressed either. I call these problems the clarification problem, the basic beauty problem and the conservation problem, respectively. I believe they are related to each other, and that depends on empirical research. Hence, addressing them is beyond the reach of the present work; however, the clarification problem should be attended to briefly, since it allows us to understand the basis of a proper notion of musical form. Before doing that, though, let us review the informal notion of form.

2. Informal Notions

Although a conceptual clarification of the idea of form is absent, there are informal ideas of musical form, based mainly on exemplification, or implicit in musical practice and philosophical discussions. I summarize those notions here.

The most common idea of musical form, perhaps the one with which most non-specialist people are familiar, refers to the large musical compositions we call symphonies, concertos, sonatas, fugues, and so on. These ways of organizing musical material have been developed historically and they are regulated by tradition. The styles

of composition are acquired and developed based on learning from previous styles or composers. I call this idea of musical form *L-form* (for large-scale).

Now, these large musical works consist of smaller structured sections of music. They constitute a more specialized class of forms that musicians use to refer to the sublevels of organization in a musical work; they include things that musicians call motifs, phrases, passages, pieces, and so on. For example, one can recognize that a certain melodic or harmonic pattern appears once and again at regular intervals during the unfolding of a musical work; these structures are called motifs. These ‘intermediate’ arrangements can also be arranged in other intermediate kinds of forms. For example, the arrangements of motifs results in passages and the arrangements of passages give us pieces. Some types of intermediate structures have special names. For example, a *binary form* consists of two rhythmic or melodic motifs **A** and **B**, organized in the simple form **AB**. A Rondo consists of a basic motif **A** and the introduction of different motifs **B**, **C**, **D**, etc., in the following fashion: **ABACADA...** All these intermediate structures, including the ones with generic names like motif, phrase, passage or pieces, and the ones with special names like binary form, sonata, rondo etc., constitute a second type of forms. I call this idea of form *I-form* (for intermediate). I-forms also depend very much on tradition, as witnessed by the fact that some of them have generic names and some others have their own proper names.

There is a third notion of form that has to do with the most elementary relations in music. Something very specific to music is music theory, a highly technical description of the rules that govern the organization of the most basic musical elements. Musical structures are thus described in their most basic elements by music theory. Musical forms can thus be interpreted as consisting, at their most basic and characteristic levels, of the items described by music theory. These kinds of structures involve, for example, the relation between two of the most basic components of music, pitch and rhythm. A melody, which is the basis of Gurney’s formalism, can be seen as a set of sounds with different pitches (musical notes) unfolding with a characteristic rhythm (which corresponds to its structure in time). Perhaps the most formalized part of music theory is harmony, which deals, among other things, with consonance, dissonance, the arrangement of notes in a scale, or their simultaneous occurrence in a chord. Most

Western music is harmonic; it consists in differently pitched sounds appearing simultaneously and unfolding rhythmically. The relation among the simultaneous pitched sounds or musical notes is not arbitrary. Sound material is musical only if it is organized according to the standard set of rules (the rules of harmony); and those rules can be understood as the set of rules that guarantee that certain sound material will be recognized as musical material by the listener. Music theory sets a framework that details the elementary relations we perceive in music. The order we perceive in musical structures can be analyzed as recurrent applications of the ‘rules’ of music theory. Music theory also describes the components and relations involved in intermediate scale musical forms. These basic level structures are embedded in intermediate I-forms which in turn are embedded in large-scale L-forms. Tradition, once again, has established how this embedding is carried out.

Now, the number of rules that constitute music theory, and especially things like harmony, is finite and those rules can be interpreted as simple manipulations of discrete elements. This fact allows the representation of musical material by a finite, yet elaborate set of symbols, and even the formalization of some musical phenomena (Riemann 1952, Estrada 1984). In a sense these basic structures are more closely linked to the idea of syntax, a system constituted by a language and rules to govern well-formedness and the construction of new musical ‘sentences’. The collection of the basic structures associated with this kind of ‘syntactic’ regulation can be seen as another type of musical form. I call this basic-level notion of form *T-form* (for music theory).

3. Clarification of Form: Basic-Beauty and Preservation

I will not address the basic-beauty and the preservation problems, but I shall examine here some aspects relevant to the clarification problem.

I believe that one of the reasons why the notion of form remains vague is that there are conceptual divisions which formalism overlooks: most formalist theories do not distinguish among their different kinds of explanations. Kivy’s formalism, for example, places a remarkable emphasis on cognitive engagement. But I think the phenomena spotted by this emphasis are not the same as those referred to by Hanslick when he

speaks about the beautiful in music, since this last notion seem to be related to mere passive contemplation. I think an explicit distinction between passive and active engagement with musical forms must be introduced: we can say that, in addition to the inherent beauty (available from mere contemplation) of some musical events, there is a second kind of musical beauty, derived not from our passive contemplation of music but from its active, cognitive processing. Thus, formalism's accounts of musical beauty can be divided into two types: passive and active. This division, I believe, must be part of a proper notion of form.

I say that an account of musical events is of a *basic type* when they are based on the fact that there are kinds of musical events that produce joy due to the contemplation of their inherent properties. For example, the sound of a skilfully played violin or a minor chord produced by an orchestra posses beauty that does not involve reference to emotions, but nor does it involve cognitive processing either. A formalist account of these events should thus be of a basic type.

I say that an account of musical phenomena is of a *cognitive type* when it states that certain kinds of intellectual activities or events in the listener produce joy. This joy is related to the effort the listeners devote to discriminating subtleties in the musical material they are listening to. For example, in the hypothesis game and the game of hide and seek in Kivy it is not the mere contemplation, but rather the cognitive activities that produce joy.

The kind of the events involved in each type of accounts is different. The basic type is concerned with contemplative perceptual events, while the cognitive type is concerned with active cognitive processing of events. Of course both kinds of musical phenomena are related to each other and the existence of this relation should be, at least, contemplated by an adequate notion of musical form, even if a full explanation of the relation is not provided.

The joy involved in the explanations of the cognitive type depends, I believe, on the joy involved in the basic type and on the *preservation* of the enjoyable properties when we make an extension from the basic to the cognitive type. In other words, in an account of how the enjoyments of basic events relates to the enjoyment of cognitive

events we must envisage that our manipulations of musical material should guarantee that the enjoyability (the beauty) is preserved when going from the basic to the cognitive events. I cannot give an account of how this beauty-conservation relation can be understood. However, concerning our clarification of the notion of form, we can simply assume that there exists some kind of conservation mechanism that links basic and cognitive events in music.

4. Clarification of Form: Non-concreteness

In addition to the basic-cognitive distinction there are features of the notion of form that derive from the ontological peculiarities of music. A very salient feature of any piece of music is that it is not a concrete object. Musical forms are abstractions of the 'concrete' features of music, so they should be even less concrete entities. According to Scruton (Scruton, 1997, pp. 3-5) causal relations are not relevant to musical phenomena. This view can be taken even further: I believe musical form is not only cause-independent, but also instance-independent. Kivy's characterization of musical experience will help us to see the reason for this development.

If a notion of musical form has to be adequate to embrace Kivy's hypothesis game and the game of hide and seek, it has to be suitable to be 'played cognitively'. The game of hide and seek involves a search for simple patterns among the complex fusion of sounds characteristic of polyphonic music. Furthermore, the melodies or motifs hidden in the mixture of sounds that constitutes music are those intended by the composer. Thus, the hidden motifs must be the same patterns every time the music is played, for they are the motifs intended by the composer; otherwise there would not be any 'seek' in the game. And we seek only because we know there is something hidden. We also know that the hidden things were hidden by someone else, the composer; otherwise the 'game' would not be a game. Thus the motives involved in the game of hide and seek must be the same every time if the performance of the work of music differs.

The problem appears when we employ a quasi-concrete notion of form based on the idea that music is a sound structure. Sound structures can be interpreted as the physical structures resulting from the sound emitted by the musical instruments when

performing a piece of music, or at least as something closely related to them. There is a tension between this view of form as sound structure and Kivy's account of music in terms of his two 'games': the games are played cognitively; the listener must seek and find the same patterns every time, as intended by the composer, even if the performance of the piece (the sounds that constitute the mixture in which we actually seek) is different every time. If musical forms were sound structures they would depend on the variation of sounds. More specifically they would depend on the actual sounds that constitute the piece in each performance. But sounds do differ in every performance; that is precisely what makes performances interesting. If sounds are different, forms are different in every performance. But this conclusion contradicts the idea that the forms we seek are those intended by the composer, regardless of the differences of performance. If the composer intended certain musical structures to be found, but the structures differ in each performance, those structures could never be found, and the hide and seek game would be pointless. Thus, if forms were concrete sound structures, the patterns hidden in the music would be different for each performance and they could never be found.

Of course, one might argue that performance variations are trivial due to the similarity of the different performances. This similarity would allow us to play the game. Then, though, we would have to postulate similarity classes as the subject matter of the game, which amounts to the introduction of a game played with abstract entities; the similarity classes. And this time we would be facing not only non-concreteness but also the abstractness of entities like classes.

I believe the games proposed by Kivy are actual elements of the aesthetic appreciation of music. My conclusion is that his notion of form needs to be revised; music is not a concrete sound structure. We already know, from the ontology of music, that form is not a physical object; it does not depend on causal relationships. Also, from the first part of the previous argument, we know it does not depend on any concrete instance of a musical piece (a musical performance). In short, musical form is instance-independent⁵⁶.

⁵⁶ We might also call this property token-independence, but I prefer 'instance-independence' to keep some distance from the standard type-token distinction, since it seems to me that the case of music is a little more complicated.

Furthering this line of thought we can remember that Scruton drew our attention to the fact that musical events have no relevant causal relations: musical events take place in an acousmatic space, not in physical space. We have seen that forms are instance-independent, now we can follow a strategy similar to Scruton's and direct our attention to the fact that musical forms are neither physical objects nor instances of sound complexes: I propose thinking of musical forms as events that take place in a more 'abstract' space, analogous to Scruton's acousmatic space.

5. Clarification of Form: Syntacticality, Plots and Patterns

In addition to characterizing music as a sound structure, Kivy presents another view of music: music has to do with internal representations (music as an intentional object) governed by syntax-like rules. Thus, Kivy's notion of music seems to have two aspects: one as sound structure and another as a 'logico-linguistic' or syntactical structure. The idea of sound structure has been discussed above. Let us now consider the syntactical aspect.

Kivy refers to the structures in music as plots without content (I call them content-deprived plots). The idea of plots without content is useful only up to certain levels of musical structure. Large and intermediate scale structures can be described as plots, but the structures described by music theory have nuances, the structure of chords or scales for example, that seem less suitable to the idea of plots. The rules of harmony in music theory do not have a narrative-like structure. Now, calling musical structures 'plots' is something of a metaphor and perhaps Kivy's characterization of music in terms of a 'pattern-regulated' syntax is more accurate.

The idea that a 'pattern-regulated' syntax governs musical forms seems to provide a more accurate description of musical structures, including those of music theory. Although interesting, this idea has a shortcoming: music, as described by music theory, is articulated in a 'two-dimensional' fashion. The 'horizontal' dimension is given by the divisions of time that constitute the rhythm of a particular piece of music. The 'vertical' dimension would be given by the different notes that constitute the musical scale. In this second dimension a complex superposition of events (musical notes, chords) occurs,

governed by rules different from those of rhythm. Furthermore, the second dimension is a dynamic one, for its structure changes when the composer modulates to a different tonality. The idea of syntax is attractive because it could account for the logical necessity that many authors attribute to music; but it must be developed to account for the multiple dimensions of music.

Musical forms, I think, are complex, multi-syntactical structures. The different syntaxes are specified by different sets of rules in music. I suggest that a reasonable way to approach this multi-syntactical feature is to see each syntax as a dimension of the space in which the musical experience occurs. Musical forms should be interpreted as objects with parameters determined by the dimension of the space in which they exist, in a fashion similar to the way we interpret mathematical spaces. Musical forms should thus be seen as abstract structures existing in a space the dimensions of which are constituted by the rules of music theory.

6. A Notion of Musical Form

Considering the features discussed above I propose here a draft notion of musical form. The reader should be aware that this proposal is only an exercise that prepared the way for my idea of aesthetic intentional object in mathematics and it is thus still fragmentary and open to revision. There are two further things that should be noted: first, I restrict myself to the case of traditional harmonic music; second, I deal with form by focusing on its 'cognitive' features.

A Musical Form complies with the following description:

A) Non-concreteness

A.1) Musical Forms are abstract structures. They are not physical events, like sound; they have no causal dependence or spatio-temporal location. They are not instances of music itself, i.e. they are not acousmatic events as proposed by Scruton.

A.2) In analogy to musical events that take place in an acousmatic space, musical forms are events that take place in a phenomenological cause-independent and instance-independent space. I have taken the term phenomenological from Scruton, but I assign it

a more restricted use; by phenomenological I merely refer to the quality of events that are experienced by the listener; in contrast to the experiences of the acousmatic space, which are musical experiences, the experiences I am referring to are ‘cognitive’ experiences.

A.3) The general properties of musical forms will thus be determined by the kind of events that are allowed to take place in this phenomenological space. We have to establish, at least informally, the features of this space.

B) *Formal Scene*: Internal Syntax

Musical Forms are events that take place in a ‘space’ analogous to the acousmatic space. I do not give a formal characterization of this ‘space’, and in order to avoid confusion with the technical meaning of the term *space* I use the term *scene*. Thus, my analogue to the *acousmatic space* could be called *formal scene*⁵⁷. Informally I characterize a formal scene as follows:

B.1) The formal scene is a ‘cognitive’ analogue to the acousmatic space, but it differs from a phenomenological auditory-space in three respects:

B.1.1) It exhibits physical-cause independence and instance independence, but it is culturally dependent; furthermore, it is culturally and historically determined. The formal scene is determined by a set of norms (several syntactical systems for rhythm, harmony and modulations) that vary according to cultural standards. By contrast, the proposed causal-independence and instance-independence are not determined by cultural standards but by perceptual mechanisms that allow us to recognize music and patterns in the music. In other words, experiencing music is something that has to do with the way a normal perceptual and cognitive system in human beings determine; but experiencing musical forms is the result of an active cognitive processing of the experienced musical material. The adaptive nature of human cognition is responsible for the plastic, cultural and historical character of musical forms.

B.1.1.1) The cultural standards for notions like harmony, tension, rest or resolution are the norms or, if you will, the ‘syntactical systems’ that organize, that give an order to, the formal scene. That leads us to:

⁵⁷ Although the term ‘space’ is possible, I reserve it for use in my notion of phenomenological space for mathematics.

B.1.2) Drawing an analogy with the idea of *ordered space* we could say that the formal scene is an ‘ordered scene’ with at least two ‘dimensions’. (The number of dimensions is also culturally determined; the two chosen dimensions presented here correspond to the traditional harmonic system of music). One of the dimensions, the ‘horizontal’ dimension, is ordered according to the syntactic rules of rhythm. The other, the ‘vertical’ dimension, is ordered according to the syntactic rules of scales, chords, and modes.

B.1.3) In addition to the two (or more) dimensions and their respective syntactical systems, there is an interdependence between them, governed by the rules of chord progressions and tonal modulation.

C) External Syntaxes

In addition to the ‘internal syntaxes’ that organize the basic dimensions of the formal scene, there are ‘external syntaxes’ that allow ‘constructing objects’ in the formal scene. I do not think we have to take the idea of ‘constructing objects’ literally; rather, we can see the ‘external syntaxes’ as rules that govern the introduction of events in the scene. Given that the forms are events in the formal scene, we can use the metaphor of building objects just as a shorthand way of naming ‘a proper introduction of events in the formal scene’. The two methods we can use to construct musical forms (to introduce proper events in the formal scene) are what I call *substructure-embedding* and *referentiality-embedding*.

C.1) Substructure-Embedding

Substructure-embedding refers to the embedding of structures already present in the scene to build new forms.

If there is no embedding in a form we call it a first-order form. If there is a single embedding we call it a second-order form. If there are two embeddings it is a third-order form. First-order forms correspond to what I called T-forms (structures governed by basic music-theoretical relations); second-order forms correspond to I-forms (intermediate scale structures; motifs, phrases, pieces, etc); and third-order forms correspond to L-forms (arrangements of whole works like symphonies). The structure of symphonies or

concerts and the arrangements of motifs inside their sections have developed historically. Harmony seems to be more stable. That is the reason why first-order forms differ from second- and third-order forms: the second and third are determined by tradition. I-forms and T-forms have been established, to certain extent, by historical tradition. The procedures of embedding must thus reflect how those forms are dealt with traditionally. Although simple recursive embedding (simply building larger structures based on smaller ones) is possible, this procedure does not reflect the actual musical forms. This is the reason why traditionally determined embedding procedures are preferred to simple recursive embedding.

The way embeddings take place obeys different rules for Second and Third Order. Although recursive embedding is possible, the traditionally established patterns of embedding will be preferred. The embedding is governed by culture-dependent 'syntaxes'.

C.1.1) For first-order forms we have the syntaxes of harmony and rhythm, already present in the structure of the formal scene. Following those syntaxes does not result in embedding, it only determines whether a certain structure is a musical form or not.

C.1.2) For second-order forms we can embed only first-order forms; we cannot use any musical event that is not already a first-order form, like a note or a chord, as part of the syntactical structure; and the embedding must follow the rules of traditional musical forms like binary forms, sonata, rondo etc. The arrangement of those forms determines the 'syntax' for second-order substructure-embedding.

C.1.3) For third-order forms we can embed only second-order forms, we cannot use any basic musical event (notes, chords, etc.) or any first-order form as part of the structure; and the embedding must follow the rules of the major musical forms, like symphonies or concertos. The arrangement of those forms determines the 'syntax' for third-order substructure-embedding.

C.2) Referentiality-Embedding

Referentiality-embedding refers to the inclusion of non-formal properties in musical structures. The non-formal properties to be included must fulfil two conditions:

C.2.1) First, they must possess referentiality (otherwise they could be treated as formal properties). Second, they must play a role in the constitution of the aesthetic value of the form in which they are being embedded. (I must recall the reader's attention to the fact that I am dealing only with the cognitive aspect of form. The most basic 'empirical' issues are addressed in the next chapter).

Emotion-qualities and basic-beauty (properties of events involved in basic-type explanations) properties are examples of properties that could be introduced as part of musical forms through referentiality-embedding.

C.2.2) The incorporation of these properties as part of a formal structure can be done in two ways:

C.2.2.1) First, the inclusion of the non-formal property P is the result of a process of substructure-embedding that has the quality of being conservative of P. For example, if embedding a *beautiful* first-order motif **A** results in a *beautiful* second-order binary form **AB**, we say that the property of being beautiful has been preserved by the embedding, and that this embedding is beauty-conservative. In this example the embedding of the first-order form **A** to produce the second-order form **AB** preserved the property of being beautiful while passing from **A** to **AB**.

C.2.2.2) Second, the inclusion of the non-formal property P is necessary to complete a certain pattern. Patterns, in this context, may be understood as generalizations of syntaxes. I focus only on the simplest and most widely used pattern: resolution. For example, if we already have two sections (second-order forms) **A** and **B** arranged as **AB** and **A** is a restful section and **B** a tense one, the resolution pattern needs a rest section in order to obtain an entire emotional-resolution pattern. In this way, we are allowed to add a third *restful* (as demanded by the resolution pattern) section **C**. We obtain the form **ABC** which completes the resolution pattern *rest-tension-rest*. Notice that **C** is not necessarily different from **A**. If we use **A** to complete the pattern we obtain **ABA**; the traditional sonata form.

Remarks

Musical forms are events occurring in the formal scene I have sketched. They are abstract structures only in the sense that they are independent of cause and instance and they have an internal organization. Their ontological status is more akin to a fictional narrative than other abstract objects, like platonic mathematical objects. In other words, musical forms are not necessarily abstract *objects* in the sense of objects actually existing with no spatio-temporal location. The notion of form presented above coheres with all of the implicit notions included in the theories surveyed in this work.

The questions about conservational and basic-beauty musical events (events involved in basic-type accounts) remain unanswered, but relevant results concerning these problems could be drawn from empirical research, especially from results of research on affective and emotive reaction.

Appendix 3: Theories and Findings on Emotions

The emotive effect of non-formal musical events has been left out of Kivy's approach. He argues that the nature of these basic events is a matter for empirical research. The arousal and expression of emotion are non-formal phenomena that are often used to explain the value and aim of art. In relation to the present project they represent a suitable complement to formalism and a natural place to look for answers about the non- and sub-formal levels of aesthetic phenomena. In the following I summarize the debate between cognitive and non-cognitive approaches to emotion, which offers a more complete setting for the background to Jenefer Robinson's theory.

1. Feeling, Behaviour and Physiological Changes

The discussion on emotions can be read as a debate between cognitive and non-cognitive views on emotion. Despite the fact that cognitive theories of emotion are the most popular, I shall start by summarizing some more basic theories. The simplest theory of emotions claims that emotions are feelings. Thus, anger is just our feeling of anger, jealousy our feeling of jealousy, love our feeling of love, etc. The problem with this theory is that there are many examples of feelings that are not emotions. Feelings associated with hunger and other bodily urges and necessities, as well as sensory perceptions like feeling cold or hot, have no associated emotion, nor anything resembling an emotion. There are no 'hungry' or 'cold' emotions, for example. The second theory, proposed in reaction to the 'feeling theory', is based on behaviourism. In the mid-twentieth century some psychologists argued that emotions should be analyzed as characteristic examples of behaviour or as dispositions to behave in a certain way. According to this view, love is not the inner feeling of love but the behaviour associated with love. The problem is that there are emotions that have no accompanying behaviour. a person could secretly experience an emotion of love, for example, and never show the characteristic accompanying behaviour of love. For that reason, some proponents of the

behavioural theory prefer to analyze emotions as *dispositions* or tendencies to behave in certain ways. Thus love is not the behaviour accompanying love but the disposition to behave in that way. There are two problems with this behavioural theory. First, there are emotions that are closely related, such as shame and embarrassment, or regret and remorse, for which the accompanying behaviour may be exactly the same. According to the behavioural theory we should consider shame and embarrassment as the same emotion. However, we can readily agree that being ashamed is not the same as being embarrassed. Behaviour or behaviour tendencies cannot discriminate precisely between closely related emotions. The second problem is that the same behaviour can sometimes be motivated by different emotions, or by a non-emotional impulse. For example, a caring behaviour may be a symptom of love, but the same behaviour may be elicited by a sense of duty. In this case the theory identifies as emotion something that is not an actual emotion.

Other theorists have noticed that in addition to behaviour there are physiological responses associated with emotions. Changes like blushing, sweating or trembling hands are all changes that arise when we experience certain emotions. William James proposed that emotions are the feeling of the physiological changes during an emotional arousal; this theory is sometimes called ‘somatic feeling’ theory. Although James’ theory led to a long line of empirical research, it has some problems: physiological changes can be associated with many causes, like physical effort or illness. In addition, long-lasting emotions, like love, may not exhibit physiological symptoms all the time; however, we would still consider them as instances of emotion.

2. Cognitive Theories of Emotion

Besides the theories based on feelings, behaviour or physiological symptoms, a fourth alternative is the ‘cognitive’ or ‘judgement’ theory of emotions. The cognitive theory of emotions is the most widely accepted and it comes in many versions. According to the theory, in a specific emotion its associated inner feelings and physiological changes are caused by the judgements made by the person. An increased heart rate could be induced by several causes, such as physical activity; the characteristic of an emotion like love is

that the changes are caused by a judgement, for example the judgement that “my beloved has arrived and that he is the darling of my heart” (Robinson 2005, p 7).

This theory explains why the same example of behaviour or the tendency to behave in a certain way is sometimes the result of an emotion and sometimes not. For example, caring behaviour towards a beloved person could be understood as the result of being in love, a state which includes the judgement that the person is a cherished person. It is still true that the same behaviour can be the result of a sense of duty towards that person. In both cases the behaviour is the same, but only the first can be understood as the result of love, since it possesses the appropriate judgement, while the latter does not.

The judgement theory also makes more precise discriminations of emotions: it explains how the same behaviour may be characteristic of two closely related emotions. For example, the behaviour associated with shame is fairly close to the one associated with embarrassment; both involve withdrawal and hiding, for example. The difference is the kind of judgements involved in each emotion. Shame is commonly understood as including the judgement that ‘I have been degraded in some way that casts doubt on my sense of self-worth’. In embarrassment I see myself, I judge I am in a socially awkward situation, but not in one that is necessarily degrading to me or that affects my sense of self-worth (Robinson, 2005, p. 8-10). The difference between remorse and regret seems to be cognitive, too. In regret I judge that something unfortunate has happened which I wish had not happened but for which I am not necessarily responsible. In remorse I judge that I have behaved morally badly, that I am responsible, and I fervently wish I had not behaved in that way.

Aristotle, Descartes, Spinoza, and Hume are early judgement theorists. The traditional philosophical investigation of emotions emphasizes looking for the judgement-content of emotions. In general terms it seems that to understand an emotion is to understand the specific judgement associated with that particular emotion. However, different authors take different approaches, often exploiting various weaknesses of the original, general approach. Robert M. Gordon, Gabriele Taylor, Robert C. Solomon, William Lyons, Martha Nussbaum, and Peter Goldie; and the psychologists Richard Lazarus, Andrew Ortony, and Phoebe Ellsworth endorse some form of judgement theory of emotion.

Robert Gordon uses cognitive science as a framework to deal with emotions. For him, the cognitive ‘aetiology’ of an emotional state determines its emotive character. Emotions are mental states identified by the particular structures of beliefs and desires that *cause* them (Gordon, 1987). Gordon treats emotions as belief-desire structures that hold causal relations to mental states and behaviour. As in the case of the behavioural theory, Gordon’s theory is too dependent on its theoretical specifics. Simple episodes of fear, for example, cannot be accounted for, since they do not include a cognitive content; especially if these episodes occur in a very short time. A further problem is that we can hold the same set of belief-desire relations that, according to the theory, should be considered as the proper cognitive state of love, for example, and even fail to experience the emotion *qua* emotion.

Robinson criticizes Gordon’s position on a similar basis; she bases her criticism on the kind of examples he offers. According to Robinson, Gordon’s examples are very “thinky” (Robinson, 2005, p. 9). They are all propositional, and they include such ‘unemotional’ examples as ‘*fear that it will rain*’ and ‘*regret that one has stepped on a pine cone*’. And it seems perfectly possible to be in an ‘epistemic’ emotional state appropriate to a certain emotion without actually *feeling* emotional at all (Robinson, 2005, p. 9). However, Gordon’s theory has a very remarkable feature; it incorporates *desires* as well as beliefs as part of the structure of an emotion.

Gabriele Taylor introduces *explanatory beliefs* in the structure of the emotions of shame, pride and guilt, as a second sub-level of beliefs in addition to an identificatory belief. An example will help to clarify this theory. Taylor analyzes pride as follows: “a person who experiences pride believes that she stands in the relation of belonging to some object (person, deed, state) which she thinks desirable in some respect...and it is because this relation holds between her and the desirable object that she believes her worth to be increased” (Taylor, 1985, p. 41)

In other words, when I am proud (in an emotional state, not as a result of a personality trait), I believe that my own worth has increased *because* a certain object, which I believe to be desirable, belongs to me. In this example there are two kinds of beliefs. First the belief that identifies the emotion, the *identificatory* belief: ‘*My worth has increased because certain object belongs to me*’. This belief constitutes the core that

distinguishes pride from other emotions, even closely related ones. Inside this belief there is a secondary belief that *explains* why I think my worth has increased, the *explanatory* belief: *This object is a very desirable one*. Taylor's theory represents a refinement of the belief-desire relations in emotions; however, it suffers from the same problems as Gordon's theory.

Robert C. Solomon addresses emotions in a different way. For him an emotion is a special kind of judgement or set of judgements that always involves a personal evaluation of the significance of an incident (Solomon, 1976, p. 178). It is an *evaluative* judgement about 'our Selves and our place in the world'. An emotion is a judgement that concerns matters that are very important to oneself, and one's interests, values, and goals. This theory recognizes that emotions have *intentionality*; they are about something. That seems to be the reason why emotions cannot be reduced to feelings, or physiological changes, or examples of behaviour. Solomon remarks on the fact that a change in the *content* of an emotion is a change of emotion itself: I cannot be angry that you insulted me if I learn that you did not in fact insult me (Robinson, 2005, p. 11). Emotions seem to have as a crucial part an evaluative judgement; we evaluate a certain situation as right or wrong according to our values; that evaluation is a necessary part of the emotions we experience. Another very remarkable fact is that when we argue with someone else about emotions, we can only argue about *evaluative* judgements, about the evaluative content of emotions. When we argue with someone about emotions, we try to convince him that he is right or wrong in making a particular evaluative judgement. For example, if I argue with someone about the fact that he is angry with me, I do so because I think he is mistakenly angry with me, and the reason is very likely because I think he is making a mistaken judgement about me. He might think I broke his new chair, for instance. Then our argument would be about whether or not he is right in believing that I broke the chair. I would try to convince him I did not break the chair. In this attempt to change his emotional state (which is usually the point of arguing about emotions) I try to change the evaluation which is causing his anger. It would be very strange if I were to argue in order to modify his accelerated heart-rate, or his blushing. This testifies to the importance of evaluative judgements in emotions.

William Lyons defines emotion as “a physiologically abnormal state caused by the subject of that state’s evaluation of his or her situation” (Robinson 2005, p. 12). For Lyon, emotion is not *identified* with a judgement but with the physiological state caused by the judgement; emotions are *caused by* ‘cognitive evaluations’ of the situation; different emotions are distinguished by their characteristic evaluations. An emotion is based on knowledge or belief about properties. For example, love is based on the (loveable) properties the beloved is believed to have; anger is based on an evaluative belief that someone has done me wrong. Lyon’s approach summarizes a relevant characteristic of cognitive theories: in all of them the judgements that are involved in emotion are evaluative judgements about a situation in terms of one’s own wants, wishes, values, interests, and goals. They are evaluations of the personal significance of something going on in the (external or internal) environment.

Richard Lazarus is among the psychologists who defend a cognitive theory of emotions. For him the relevant ‘judgement’ that forms the ‘core’ of an emotion is always an appraisal of the significance of the person-environment relationship. Lazarus calls the emotive assessment of certain situation a *core relational theme*. These assessments or appraisals are evaluations that the organism makes concerning the potential harm or benefit of a situation, or an ‘adaptational encounter’. The specificity of the relational harm or benefit underlies each specific kind of emotion, or, rather, a family of emotions (Lazarus, 1991).

Lazarus claims that the appraisal is both a necessary and sufficient condition for emotion to occur. In addition, he uses the ‘core relational themes’ to develop a typology of emotions. Each core relational theme identifies the evaluations necessary for a ‘basic emotion’ type. The basic emotions he identifies are anger, anxiety, fright, guilt, shame, sadness, envy, jealousy, disgust, happiness, pride, relief, hope, love and compassion. Each basic emotion possesses its distinctive core relational theme, which is a specific appraisal of the potential harm or benefit that arise when the organism has to adapt to a new situation.

Like Lazarus, the psychologist Andrew Ortony considers that appraisals of harm and benefit play a central role in emotion. He has a different research strategy but he is also concerned with emotion typologies. Ortony distinguishes three classes of emotions,

based on the different kinds of appraisal that produce them: appraisals rooted in goals, appraisals rooted in standards and norms, and appraisals grounded in tastes and attitudes. Ortony also identifies different types of ‘valence’ for the emotion classes. These valences serve to model the nuances of emotion based on the differences in the way the appraisal is made. Ortony’s valences are: the desirability or undesirability of events, the praiseworthiness or blameworthiness of agents, and the appeal or lack of appeal in objects. Finally, Ortony speculates that the role of emotions is to focus the attention of the organism on the changes in the environment that can affect it: emotions *represent* in a conscious and insistent way (through distinctive feelings and cognitions) the personally significant aspects of situations from our subjective perspective (Ortony, 1988).

3. Criticisms of Cognitive Theories

According to the judgement or cognitive theories, a judgement is a necessary condition for emotion. However, a general problem with these theories is that a judgement by itself does not seem to be a *sufficient* condition. We can make the relevant judgement, as specified by a certain theory, and fail to feel the corresponding emotion. For example, I can judge I am in a very awkward social situation. According to the theory I should be experiencing embarrassment. However if I am in a very good mood, for example, I can feel amusement instead of embarrassment, or simply fail to feel anything new and remain in my initial pre-embarrassment state. I can make a dispassionate evaluative judgement, about a potentially harmful situation, for example, but it is plausible that the appropriate emotion associated with that appraisal – fear – fails to appear; maybe because I am in a euphoric state or very depressed. There is nothing in the *evaluation* component that necessarily adds a passionate response to a non-evaluative judgement. By bringing evaluations into the judgements that identify emotions we certainly improve our insight into the kind of mental contents that usually appear in different emotions, but the ‘passion’, the emotional part of an emotion, remains absent.

Even if we add *explanatory* judgements the situation still does not improve. Take embarrassment, for example: I can judge I am in an exceedingly awkward situation and I can judge that being in an awkward situation is undesirable. But I can still react

dispassionately to both judgements and, thus, miss the emotional part of the supposed emotion.

A solution to this problem has been suggested by Solomon. He suggests that judgements involved in emotions are a kind of ‘urgent’ judgements; emotions are ‘self-involved and relatively *intense* evaluative judgements’. The judgements and objects that constitute our emotions are those which are specially important or meaningful to us. Emotions are responses to difficult situations, but they are ‘intense’ judgements in response to difficult situations (Robinson, 2005, p. 15). Robinson complains that calling judgements ‘intense’ is odd; and calling them ‘urgent’ is problematic. I would add that changing the *quality* of judgements does not solve the problem unless it explains or implies the *emotional qualities* of emotion. Neither the intensity nor the urgency of a judgement changes it from a set of mental contents to a passionate experience.

The characteristic content of judgements proposed by judgement theorists, an evaluative content directed to something, together with the introduction of causal relations in the organism that experience the emotion suggested by psychologists like Taylor, led Robinson to propose a different view. Robinson remarks that the evaluations associated with emotions are evaluations of the environment that assess different aspects of it: certain circumstances could be seen as a threat or as an offence, as friendly or hostile, for example. Robinson argues that emotions are not exclusively human; she suggests that emotions are ways in which organisms interact with their environments. Furthermore, she thinks that it is not accurate to picture emotions as private events happening in our minds; rather, she suggests that emotions are *interactions* with the environment. She proposes a change in emphasis from the intentionality of emotions to their causal relation to the environment.

In Robinson’s proposal, emotional states are not directed towards propositions; rather, they are provoked by the environment which in turn is perceived under a particular aspect, as threatening, amiable, offensive, etc. When humans or other species interact with the environment they do so in several ways; sometimes, emotion is one of those ways, distinguished by the fact that in emotional interactions the environment is viewed in terms of its effect on the wants, interests, and goals of the organism.

4. Non-Cognitive Theories

Before surveying the empirical research on emotions we must review some of the critical views on judgement theories. An important criticism is that the content of the evaluations involved in emotions is not necessarily believed or judged to be true. A stronger version of this criticism denies any propositional content to emotions; an emotion is experienced more like perceiving or paying attention to something in a certain way.

Patricia Greenspan, for example, criticizes judgement theories, arguing that emotions cannot be judgements because they have different rationality criteria. She argues that emotions, unlike judgements or beliefs, have formal logical properties that allow internal inconsistencies or contradictions inside the system of beliefs. Greenspan suggests that the experience of ‘mixed feelings’ shows the different rationality of emotions. For example, if a friend of mine, whom I love, won a prize that I wanted to win, I would feel happy that he won the prize but unhappy I did not. The judgement that my friend won a prize would then be qualified as something good *and* bad, as something simultaneously desirable *and* undesirable. Greenspan claims that people do not need to sum up or qualify their judgements to be rational, at least in the case of emotions. It is perfectly rational for me to *experience* happiness and unhappiness caused by my friend’s winning the prize of which I dreamed. However, it is not rational for me to hold the contradictory judgements that my friend’s winning the prize is something desirable and that my friend’s winning the prize is undesirable; thus, emotion must be something else.

Greenspan notes a second problem with the idea of emotions as judgements: the relation between rational judgements and evidence is different from the relation between emotions and evidence. An impartial judgement has to be assessed for rationality based on all the available evidence. Emotions, by contrast, seem to be based only in a subset of the available evidence: they are focused only on the information or features that are meaningful or important to us (Robinson 2005, pp. 20-23). Thus, the simultaneous happiness and unhappiness that my friend won the prize I wanted to win can be understood by considering that they are based on different, inconsistent subsets of evidence. This interpretation of emotions seems to relate emotions more to desires than

judgements, because simultaneous emotions can be explained as being responses to simultaneous wants: I want to win the prize, but I also want my friend to be happy.

In her later work Greenspan argues that the evaluations involved in emotions should better be described as similar to '*thoughts of*', or propositional attitudes, rather than beliefs or judgements. Emotions are states of comfort or discomfort directed towards evaluative thoughts. Instead of evaluating that a certain object is dangerous, for example, we can merely have a '*thought of danger*'. In addition, these evaluative propositions may be merely entertained and not necessarily endorsed. Irrational fears are examples of situations where Greenspan's claims seem to work very well. In a situation of irrational fear it is not always obvious that our 'normal' thinking is at work. Our emotional response seems to be rather automatic, like a reflex, with very little propositional content or none at all.

Greenspan's theory certainly clarifies some aspects of emotions that the judgement theories do not, but it still has some weaknesses. For example, suppose I am in fear; I merely have a '*thought of danger*' rather than a more complex judgement or evaluation that I am in danger. I can understand that certain aspects of my *behaviour* in response to danger are at least partially explained by the fact that I am seeing myself in danger and that something must *be done* to avoid the danger. The simple '*thought of danger*', however, is too vague and it is certainly difficult to infer from it a specific behaviour. In other words a mere *thought of danger* does not seem sufficient to explain the behaviour associated with fear. The same problem is present when we think of physiological changes instead of behaviour. I can have thoughts of danger and, as in the case of judgements, fail to feel or have the physiological symptoms of fear. The problem with this theory is that it keeps directing emotions towards propositions, and thus it shares some of the problems of judgement theories: it is not adequate to explain the 'passionate' features of emotion.

Amélie Rorty rejects any propositional content in emotions, either 'believed' contents or 'held-in-mind' contents. To support this claim, Rorty draws our attention to the fact that emotions sometimes persist even though the corresponding judgement or belief has vanished or has been rejected. For example, if I am angry because a known person has stolen my car and I discover it is actually false that that person stole my car, it

is correct to say, according to the judgement theory, that I am no longer angry. However, I may still be angry with that person, perhaps irrationally, and simply look for some other 'reason' to justify it. Alternatively, I may remain angry but my anger may be transferred to another person (Robinson 2005, p. 23).

Rorty suggests that rather than judgements we should look for an alternative intentional component of emotions. She proposes that the intentional components of emotions take the form of evaluative *interpretations* which focus on certain aspects of the environment. She calls these evaluative interpretations *patterns of intentional salience*. Instead of judging a certain situation as threatening, for example, the individual *sees*, construes, the situations as a threat. Even if the person does not believe that a certain situation is dangerous, for example, this person cannot avoid *seeing* it, interpreting it, as dangerous. One advantage of this suggestion is that the *seeing-as*, unlike judgements, can be exported to emotional episodes in non-human animals.

Robert Kraut also emphasizes the relationship between emotion and perception, rather than judgements or evaluations. Kraut suggests that a better model for emotion would be perception rather than belief or cognition. For example, love should not be understood as grounded on a judgement or set of judgements about the beloved person, but rather as perception directed at a particular individual. The intentionality of emotion does not have to be propositional intentionality or propositional content. Kraut accepts that emotions are based on some kind of evaluation, but he reminds us that perceptions themselves, just as beliefs, are value-laden. (Robinson, 2005, p. 25)

The discussion above can be summarized by drawing the following conclusion, as suggested by Robinson. First, emotions are not simply judgements; however, some kind of appraisal or evaluation seems to be necessary. Second, there is a connection between emotions and our interests, wants, wishes, values, and goals. Third, there is disagreement about what kinds of evaluations are necessary for emotion: evaluations in emotions can be embodied in a belief, an unconscious appraisal, a merely *thought-of*, a perception or *seeing-as*, or in something that looks like a reflex action. Finally, a general problem is the lack of an account of the psychological and physiological changes associated with emotions.

5. Experimental Psychology Results

No contemporary image of emotion would be complete without the results of experimental psychological research on emotional responses. Studies of physiological changes, for example, have been used to characterize a range of emotional states. Research on emotional responses in lower organisms seems to indicate that emotional reactions have an innate component, and they are related to adaptive behaviour to the environment. In human beings physiological, facial, and vocal changes seem to be an automatic adaptation of the organism to deal in an immediate way with new or significant situations in the environment. The results of empirical research seem to indicate that emotions are a part of a state of readiness that the organism adopts semi-automatically in order to deal with the environment. We shall see how neurophysiology offers evidence against judgement theories and in favour of an image of emotions as adaptive-response processes which consist of automatic appraisals triggering physiological responses, which are then inhibited or reinforced by cognitive appraisal.

6. Physiological Changes in Emotion

William James' theory of emotions has been one of the most influential works on psychological research of emotions in the twentieth century. James' theory of emotions is often read as claiming that emotion is the feeling or awareness of the physiological changes of emotional responses. Robinson, however, construes James as introducing, in the very early times of scientific psychology, the notion that physiological changes are essential to turn an intellectual perception into an emotional state. Evaluations are not sufficient for emotions; we also need physiological changes, especially in short-term emotional responses like fear.

Walter B. Cannon, a student of James, disagreed with James about the general character of the physiological change involved in emotions. James thought that the physiological change exhibits a different 'profile' for each emotion (that would be the reason why we experience, we feel, different 'feelings'). Cannon, by contrast, thought

that emotions involved a generalized pattern of arousal in the nervous system⁵⁸, which is also present in some non-emotional states such as pain, hunger or cold, and which he called the ‘emergency reaction’.

Many empirical results have shown that the pattern of arousal is different for different kinds of emotions. A. F. Ax (Robinson, 2005, p. 30), for example, reports that fear and anxiety involve an increased secretion of epinephrine, but anger involves increased secretion of norepinephrine. Furthermore, tranquil or calm emotions – calm amusement, for example – reflect a dominance of the parasympathetic instead of the sympathetic nervous system.

The evidence for claiming that there are specific patterns of arousal for different emotions is still not completely persuasive. However, some specific physiological changes – skin conductance or heart rate, for example – seem to exhibit a fairly reliable relation to patterns of arousal depending on the specific emotional response. Robert Levenson (Levenson, 1994, p. 255) has identified four (plus a fifth) reliable physiological differences among the emotions of anger, disgust, fear and sadness:

- (a) anger produces a larger increase in heart rate than disgust; (b) fear produces a larger increase in heart rate than disgust; (c) sadness produces a larger increase in heart rate than disgust
- (d) anger produces a larger increase in finger temperature than fear... The possible fifth difference is that sadness produces greater peripheral vascular dilatation and greater speeding of blood to the periphery than the other negative emotions. (Levenson, 1994, p. 255)

An important feature of these results is that the differences are relative to other emotions, there is no physiological ‘signature’ for sadness, but only sadness differences relative to other emotions.

The fact that there are no specific patterns for different emotions is reinforced by the fact that physiological changes exhibit variations for the same emotions under different circumstances. Different instances of the same emotion can be identified; for example, there is calm joy and active joy, or rigid and mobile fear.

In view of all these facts, Nico Frijda (Frijda, 1986, pp. 133-135) takes a different approach. For him the different patterns of physiological activity in emotion correspond to the action requirements of the situation that triggered the emotion. In that sense, the

⁵⁸ More specifically the sympathetic system.

physiological changes observed in an emotional response are associated with particular states of action-readiness, rather than with particular emotions. This explains why the responses differ from one study to another and from one 'kind' of emotion to another: physiological response patterns correspond to the functional requirements of dealing with the environment rather than to different emotions. Physiological responses, in this view, are functional preparations of active responses in order to deal with the environment; they are not identificatory 'signatures' of emotions, but rather responses elicited due to the need to deal with a certain event. Even if some distinction among basic emotions can be made, physiological changes do not provide us with specific profiles that would allow us to identify a particular emotion. However, we shall see that facial expressions exhibit a remarkable ability to 'tell' the emotion associated with them.

7. Facial Expression

Changes in the facial musculature are some of the most remarkable instances of physiological changes and perhaps the ones that are closest to providing an 'identification profile' for emotions.

Paul Ekman was the pioneer in the study of facial expression (Robinson, 2005, p. 32-33) in emotions. He has concluded that there are universal facial expressions for anger (Ekman, 2003), fear, enjoyment, sadness and disgust; and, more recently, also for surprise and contempt. Ekman does not identify fixed expressions for specific emotions but rather expressions for *families* of emotions which exhibit differences depending on the strength and type of emotion inside the emotion-family. Emotion-families share particular muscular patterns by means of which they can be reliably distinguished.

The claim about the universality of facial expression of emotions is supported by a high degree of agreement in the labelling of expression across different cultures (both literate and preliterate), across different social contexts; and both in deliberate and spontaneous expressions. However, there is a problem with the claim about the universality of facial expressions: sometimes the same facial expression seems to be the result of different emotions in different cultures. Ekman explains this phenomenon by arguing that it is not the relation between an emotion and its facial expression which is being determined by the culture, but rather the 'suitability' of showing certain emotions

in certain social contexts. He claims there are simply different ‘display rules’ for different emotions in each culture, and he supports his claims with experimental results. For example, a movie was shown to two different groups of people coming from America and Japan; their facial responses were then observed. In the second phase of the experiment the subjects were told they would be observed during the movie. The results showed that Western people remained expressive in both cases, but people from Japan tended to hide their expression when they were told they were observed. Other experiments have shown that even when a display rule prevents us from showing a certain emotion, there are certain facial traits that ‘leak’ through our facial expression. Ekman calls these leakages ‘micro-expressions’. A related result showed significant differences between deliberate, voluntary expressions and spontaneous ones. Finally, studies of injured subjects conducted by Antonio Damasio (Damasio, 1994, 138) seem to indicate that some facial expressions of emotions use different neural pathways depending on whether they are produced voluntarily or spontaneously (Ekman 1994, 2001).

A very remarkable result of Ekman’s research shows that facial expressions can cause physiological changes, especially in the autonomic nervous system: when certain subjects were asked to configure certain facial musculature patterns, without any knowledge of the specific expression intended, they also showed the physiological changes in skin temperature and heart rate distinctive of the emotion associated with the facial expression. In other words: to some extent, facial expressions can induce the emotion of which they are an expression.

These results led Ekman to believe that there may be what he calls ‘affect programs’, which are sets of responses that are complex, coordinated and automated, and that consist of ‘autonomically induced systems of reactions, including autonomic, facial and vocal expressions’.

These results offer scientific support to the idea that there are emotional states that can be evoked without any evaluation – against the main claims of judgement theorists. Now, physiological-reductionist theories, like James’ somatic feeling theory, sound very simplistic in comparison with judgement theories. However, the conceptual shift and empirical research started by James showed that emotions are, in part, a set of physiological changes triggered by an assessment of the environmental circumstances,

rather than a judgement. The physiological changes can be, as Frijda suggests, part of a state of readiness that the organism achieves in order to deal with the circumstances.

8. Affect and Cognition

Experimental psychology offers further evidence against judgement theories. Developmental psychology offers results that indicate that human beings can experience emotion without the cognitive evaluation judgement theories regard as necessary for emotion. John Watson, the founder of behaviourism, reported that newborn babies, who cannot make complex judgements, universally exhibit reactions like rage, fear and pleasure under certain circumstances. Alan Sroufe (Robinson, 2005, p. 37-38) studied the development of various emotion systems and found that they are linked to the development of higher cognition; however, a basic emotional reaction is already present in newborn children. Klaus Scherer furthered this research, claiming that emotions depend on what he calls 'stimulus evaluation checks'. He proposes they may be independent, at least in part, of higher cortical⁵⁹ functions. The most elementary reactions present at birth are the check for novelty and the check for intrinsic pleasantness and unpleasantness. Both Sroufe and Scherer conclude that the cases of 'full' emotions have as their precondition the primitive emotions found in newborn children. Similar primitive emotion systems are present in other species, like dogs or apes; that is also evidence of the existence of emotions without the involvement of highly cognitive judgement.

The work of the psychologist Robert Zajonc also contributes to questioning the role of judgements in emotion (Robinson, 2005, pp. 38-40). He argued that, at least in some cases, emotion can occur without prior cognitive assessment or appraisal. The evidence for this claim comes from four main sources:

1) The *mere exposure effect* experiments. These established that people *unconsciously* prefer stimuli to which they have previously been exposed. For example, in one experiment a series of polygons were shown rapidly to the subjects, so they had no time to recognize what they were seeing. When the subjects were asked to say which

⁵⁹ The adjective 'cortical' refers to matters related to brain cortex, the outermost layer of tissue in the brain. The cortex is responsible for most highly cognitive functions, like abstract reasoning.

polygons were shown, they failed; their answers were close to chance, showing they were unable to discriminate the stimuli consciously. In contrast, when they were asked about which polygons they preferred, they chose the ‘familiar’ polygons rather than the ‘new’, previously not shown, unfamiliar ones. These results suggest the existence of a capacity for making affective discriminations without extensive participation of the cognitive system.

2) The studies of so-called *non-conscious affective priming*, which refers to the induction of an affective reaction to a neutral non-emotional stimulus, like a Chinese ideogram, by priming the stimulus with a second affective stimulus, like pictures of anger or happy faces. In the experiments the subjects were shown pairs of such images. The pairs were shown rapidly so the subjects had no time for conscious discrimination. When asked, the subjects preferred the stimuli primed with positive affective stimuli to those primed with negative affective stimuli.

3) The ‘*subception*’ experiments by Richard Lazarus and R. A. McCleary (Lazarus, 1991, 155-156). Subception was the name Lazarus gave to a kind of perception that occurs below awareness, a phenomenon observed in some of his experiments. In the experiments the subjects were conditioned, through electric shocks, to respond to five out of ten strings of nonsense letters. When the strings were later presented at a speed that prevented conscious recognition, the subjects were unable to identify the strings; however, they showed changes in skin conductance and heart rate when the string presented was one of the electric-shock-conditioned strings. The subjects did not know whether or not the string was a conditioned one, but their skin did.

4) Several experiments by other researchers showing that cognitive activity is not always necessary for affective responses. For example, John Garcia and K. W. Rusiniak (Robinson, 2005, pp. 40-41) showed that disgust responses can be conditioned in anesthetized rats. In the experiment rats were exposed to a new flavour using flavoured water: fifteen minutes later, with the rats now anesthetized, a second nauseating substance was administered. The rats eventually developed an aversion to the flavoured water even if they had no means to make a ‘cognitive’ association between both stimuli (Robinson, 2005, pp. 39-41).

9. Affective Appraisals

What all these results show is that there are instances of affective responses in which highly cognitive activity plays a minor role or none at all. However, Richard Lazarus construes these results more subtly. For him emotion is always an assessment, an appraisal, of the significance of the person-environment relationship. Emotions, in this theory, are the reactions of the organism to the matters that urgently concern its well-being. The empirical research indicates that the organism reacts very promptly to a stimulus that can be *assessed* as affecting its goals and needs or, in the case of human beings, its values, tastes and attitudes. Now, for an appraisal theorist there are two ways to deal with the results that deny cognition in affective responses. The first is just to stick to the notion of emotion as having a judgement or appraisal as its core. In that case, most of the phenomena observed in the experiments reported above would not be considered genuine cases of emotion. Lazarus embraced this way at first, but in his later work he took a different approach, which is also the second alternative for appraisal theories: he thinks that in addition to ‘full’ judgements, there is a second kind of appraisals that does not involve higher cognitive processing but rather the autonomic nervous system. Thus, he proposes two different modes (perhaps systems) of appraisals: one conscious, deliberate and under volitional control, the other automatic, unconscious and uncontrollable. The unconscious appraisal would be the appraisal system that is at work in the subception experiments, for example.

This last way of thinking is closer to the views of researchers, like Zajonc, who claim that affect can precede cognition. For Zajonc, the affective response involves appraisals that occur rapidly and automatically; among them are appraisals such as liking, disliking, disgust and threat. An explanation of these automatic appraisal systems is that they are advantageous because they take place more rapidly than higher cognitive assessments. In environmental and social contexts it is usually very useful to have an organism that recognizes and responds quickly to any change that can involve its well-being, rather than one that stops to reflect but compromises its own integrity by doing so. This view of emotions coheres with Frijda’s view on physiological changes in emotion: if emotions are adaptive responses, the response is even more efficient if it includes a

preparation of the organism to deal with environment. Physiological changes are part of this preparation.

Among the physiological changes involved in emotions we have seen that muscular configurations seem to play a role. In this respect, it has been proposed that emotions can serve as signalling devices. Zajonc and Ekman have noticed that the motor activity of an organism can serve as signals to other members of the species, signals that indicate one is in a particular state. In social situations, information about the state of other organisms is important to determine individual or collective behaviour. The ability to recognize an animal's angry attitude can help another animal of the same species to avoid confrontation or prepare for a fight. In addition, some body signals are already responses to an external stimulus; for example, looking humble, by sending a 'peace signal', can help the threatened animal avoid confrontation.

The idea of emotion-as-signal has been developed by Keith Oatley. He contrasts what he calls 'semantic messages', which have propositional content, with 'control messages', which do not need to be parsed or interpreted (Oatley, 1992, p.53) but which function merely to activate particular cognitive or motor systems. For him, emotion signals are very like control messages. The role of these 'emotion control systems' is to focus attention on some situation or event in the environment which is crucial to our needs or concerns, and they keep the attention focused as long as the situation prevails. A problem with this view is that there are many everyday situations in which selective attention occurs, but they are not instances of emotional arousal. A building inspector, for example, has to focus his attention on certain aspects of a building, and this certainly has to do with his concerns. In addition, his attention needs to be kept focused during the entire inspection; however, it would be very difficult to interpret this as an emotional episode. The problem with this model is that physiological changes do not figure in it.

A more accurate model is one in which these emotion control signals always result in physiological changes (which are probably part of the same adaptive response), which includes the appropriate facial and muscular responses.

In addition to these traditional problems for judgement theories, there is a conceptual problem. The solutions proposed by Lazarus and Oatley consider cognition as consisting of two systems of appraisal; this implies the incorporation of a second

subconscious system of appraisal into the notion of cognition. This would lead us directly into a difficult debate: what is cognition?

Robinson, who is a non-judgement theorist, rather than fuelling the cognition debate, prefers a compromise in this respect. She accepts that emotion includes a system of primitive or basic *non-cognitive* appraisal, independent of higher cognition, that evaluates changes in the environment in a very primitive way; making rapid assessments of a situation as familiar or unknown, good or bad, beneficial or harmful, etc. Questions concerning whether these assessments or appraisals are for or against judgement theories are left out the discussion because, as many results show, cognition, either 'full' or in the form of basic appraisals, is not enough for emotion. Instead, it is the relation between this appraisal and physiological responses which seems to determine whether or not a response is an instance of emotional arousal. Later we shall see that 'full' cognition can play a role in other theories of emotion.

The solution of incorporating 'affective' appraisals (as Jenefer Robinson calls the non-cognitive appraisal) into emotion has some advantages: it gives a common theoretical framework to research on humans and lower species. It can also explain some puzzling phenomena about emotion. For example, that we can have emotional attitudes that contradict our cognitive attitudes: I can be afraid of flying even if I am certain that flying is safe. The reason is that the appraisal which is triggering my emotional reaction is independent of my higher cognitive attitudes. And that appraisal is also a very basic one, it is incapable of making subtle distinctions, it just sees the situation as bad or frightening.

The introduction of affective appraisals is an important step towards improving the notion of emotion and it is also a first step to incorporating cognition in a more accurate image of emotions.

The coupling of affective appraisals and physiological changes also provides explanatory advantages. Emotions function as rapid signalling devices playing at least two important roles for the organism. First, they provide a rapid, adaptive response to potential situations that matter to the organism and, by triggering physiological, muscular and facial changes, they prepare the organism for a further response. At the same time, the visible changes serve as signals to other members of the species, providing them with

information about the subject's emotional state. Second, some physiological changes function as 'feedback' to the system (that is why there is a somatic feeling associated with emotion), reinforcing the appraisal and helping to keep attention focused on the event.

Appendix 4: More on Expression

The idea that art is expression in Collingwood's romantic sense was widely accepted in the first half of the twentieth century. However, by mid-century the situation had changed, as many philosophers wishing to distance themselves from the idealist background of romantic theories spotted some of their relevant problems. Here I summarize Robinson's discussion of the ideas of two of those authors which serve as a background for her theory of expression.

1. Hospers and Tormey

Two of the authors who debated traditional idealist expression theories were John Hospers in his 1955 article, *The Concept of Expression*, and Alan Tormey in his 1971 book, *The Concept of Expression* [sic].

Hospers describes three ways of thinking of expression (Robinson, 2005, 239-240): (1) expression as a process undergone by the artist, (2) expression as the evocation of emotion in an audience, and (3) expression as communication between artist and audience, which is a compromise between 1 and 2. Hospers thinks that the traditional theory of expression falls under variety 1: expression would be an expressive process undergone by the artist; an expressive work of art would be the final result of such a process. It is not difficult to find counter-examples for each of the three approaches. For example, a musician can be involved in a *bona fide* process of expression, and even so his work can fail to express what it was meant to express due to a poor performance of his work. In this case the type 1 theory would fail to exclude a clear non-example of expression. A world cup football match certainly evokes strong emotions in its audience, but we do not consider a Netherlands-Germany football match a *bona fide* instance of artistic expression. A type 2 theory would fail to exclude this non-example of expression. Finally, I can take my audience to see the same beautiful sunset that profoundly moved me a minute ago and succeed in sharing my joy with them. But that is not an artistic expression. Type 3 theories cannot exclude non-examples of expression like these.

Hospers thus rejects the three varieties of theories of expression and proposes a different theory of expression, 'expression as a property of the work of art'. It is neither the artist nor the audience that matters here; it is the work of art itself. Being expressive is just a property of works of art⁶⁰. This idea is more clearly presented by Tormey.

Alan Tormey (Tormey, 1971, 103-106) interprets expression as a property of the works of art (or other objects). He presents a notion of expressive quality: expressive properties, such as 'gloomy' or 'nostalgic', are those properties of artworks (or natural objects) whose names also designate intentional states of persons.

Tormey is more rigorous in his criticism of traditional expression theory and in his proposal of expression as a property. Instead of three separate theories, Tormey sketches and criticizes a generic Expression Theory (E-T) which, he claims, summarizes the characteristic features of the traditional expression theory:

(E-T) If art object O has expressive quality Q, then there was a prior activity C of the artist A such that in doing C, A expressed his F for X by imparting Q to O (where F is a feeling state and Q is the qualitative analogue of F) (Tormey, 1971, p. 103)

Tormey argues that the fundamental mistake in the Expression Theory is its assumption that the existence of expressive qualities in a work of art implies a prior act of expression. In other words, we have examples of expression that are not covered by the theory. For example, in music, a minor chord is usually thought to sound sad, while a major chord sounds cheerful. Now, a progression of two minor chords can express sadness as a result of the nature of the chords, regardless of whether the performer of the music intended so or not.

Tormey proposed an alternative view of expression, similar to Hospers' but further developed. According to Tormey, expression is characterized by two things. First, expressions are always expressions of intentional states, of states that have intentional objects. What can be expressed are only internal states that are directed towards or about something. Thus, only things like emotions, thoughts, or propositions can be expressed. Second, expressions always warrant certain kinds of inferences. In general, 'A is expressing ϕ ' implies that 'A is (or has) ϕ ', ϕ being an intentional state or, in our case, an

⁶⁰ The coincidences with Kivy's ideas are obvious, but I do not explore them here.

emotion or the name of an emotion. In this respect, the failure of E-T can be seen from a conceptual point of view. Tormey criticizes E-T by claiming it fails to distinguish between being an ‘expression of φ ’ and being a ‘ φ expression’; φ stands for some psychological state in the artist, poet, or composer.

Tormey spots a second problem with the traditional expression theory. It is true that works of art can reflect personality traits of the person that made them and in that sense they express something about its creator: however, the same can be said about any other product of a human being. The problem is that in these expressions there is nothing that warrants their aesthetic relevance. Furthermore, the same can be said of expressive qualities in a work of art, because the relation between expressive qualities and the artist’s emotional state is always contingent. Tormey uses Nielsen’s 6th symphony to exemplify this problem.

The second movement of Carl Nielsen’s Sixth Symphony (1925) is, according to some musical critics, an expression of his bitterness and disappointment due to his failure to reach an international audience. However, Tormey points out, when one listens to the music it sounds very different. Instead of bitter or ‘disappointing’, it sounds playful, humorous, or ‘buffoonish’ (Tormey, 118-119). Tormey concludes that the artists do not express themselves in their works in a sense that is intelligible, consistent or aesthetically relevant. The expressiveness of an artwork depends on its expressive features, not in the artist’s intention to express himself. Expressive qualities are emergent properties that are based on simpler, non-expressive properties. For example, the playful character of a piece of music depends on the kind of chords and harmonic progression as well as its tempo and instrumentation; no matter how miserable I feel, if I play a major chord on the piano it will not sound sad.

In Tormey’s view, to summarize, traditional expression theory is either false or trivial. It is false because we can infer neither emotional states nor acts of expression from a work’s expressive features. It is trivial because it fails to characterize art: expression does not make a work of art any different from any other human product; in other words, it fails to exclude non-artistic examples of expression.

Appendix 5: A Notion of Form Revisited

In view of the theories about emotions and expression, an improvement to my notion of musical form can be achieved. I propose a revision of the notion presented in Appendix 3. The discussion of the importance of context as an element that helps some properties to be manifested can be incorporated in a semi-formal way into our notion of form.

Briefly, a musical form is an intentional object which is defined by structural properties and rules that govern their interrelation, as well as how they combine to result in new structures. These rules constitute a kind of ‘syntax’, since they are rules of formation and combination; they determine which objects are well-formed forms and how to construct well-formed forms out of our basic vocabulary. The basic syntax of music is harmony. However, musical forms have different rules of formation depending on the level of construction. Harmony tells us how to construct melodies, chords or motifs out of pitched sounds; but if we want to construct a sonata, we need in addition the rules that tell us how to construct a sonata out of musical motifs. And if we want to construct a symphony we need a different set of rules that tells us how to construct a symphony out of pieces like sonatas or other intermediate arrangements. We have different syntaxes working at different levels of construction. I have called these different syntaxes the dimensions of a formal scene in which our intentional object exists: musical forms exist in a ‘space’ with multiple syntactical dimensions.

Now, according to Robinson (Robinson, 2005, p. 249), extra-musical knowledge allows certain properties to be manifested. This phenomenon can be modelled as a class of rules that govern the introduction of properties that are not visible within our scene or ‘space’. The most important characteristic of extra-musical context is that it must be relevant to the manifestation of an expressive property. This characteristic can be modelled as a class of rules for the introduction and transformation of non-formal properties. I call these rules meta-intentional rules, since they add contents that are not visible within our original ‘space’. For example, historical or biographical knowledge are not governed by the rules of harmony and thus they are not ‘visible’ within a purely

formalist experience of music, but they can help some expressive properties to be manifested, as the Nielsen example shows.

Meta-intentional rules are not arbitrary; rather, they are constrained by two conditions: first they must be aesthetically-conservative; that is, they cannot contradict the internal syntaxes and they cannot change any of the properties that are considered to be responsible for the aesthetic value of musical forms. For example, they cannot introduce non-pitched sounds or dissonant chords because that is against the vocabulary and rules of harmony. Second, they must help us seek, or be relevant to seeking other kinds of *patterns* in music (the most important of such patterns is resolution). For example, the fact that Nielsen's Sixth Symphony (a very playful and humorous sounding piece of music) was written during a very disappointing and bitter period of Nielsen's life is not a *formal* property of the symphony. It is not something we notice when listening to the music; the symphony sounds rather playful and humorous. However, including this property as part of our object of attention helps us find new patterns of expression: Nielsen is expressing sardonic bitterness and his despite for critics. These properties would otherwise be 'invisible' to us.