Received: 06 June 2024
Accepted: 20 June 2025
Published: 28 August 2025

Investigating the Fundamental Base of Emotion Science

Juan R. Loaiza – Universidad Alberto Hurtado, Chile, jloaiza@uahurtado.cl

Abstract

How should we investigate folk emotion concepts for the purposes of anchoring scientific emotion concepts? In this article, I expand on Mun's ideas on what she calls the fundamental base for interdisciplinary inquiry in the science of emotion. I argue that first-personal experiences should not be identified with the *explananda* of emotion science. This is because these experiences do not provide a theoretically neutral and intersubjectively accessible ground on which to anchor scientific concepts of emotions. Instead, I propose a pragmatic account of how to investigate the fundamental base, drawing on Haslanger's distinction between manifest and operative concepts, as well as some of her views on social constructionism and semantic externalism. This pragmatic account covers the relevant sources of evidence considered by Mun, but also calls for an investigation of cultural variations in emotion concepts, scripts, and norms. The upshot of the pragmatic account is a more expansive research programme that maintains the interdisciplinary spirit of emotion science.

Keywords: folk emotion concepts, first-person experience, externalism, social construction

What should emotion science explain? Emotions, of course. But how are emotions to be individuated and identified for the purposes of emotion science? Answering this question is not an easy task. On one view, emotions should be individuated and identified through ordinary intuitions. This is what Mun (2021) calls the "*fundamental base for interdisciplinary inquiry in the science of emotion*." But how should we identify those ordinary intuitions? In other words, how should we characterise that fundamental base for emotion science?

In this paper, I argue for an account of the fundamental base for emotion science, and I propose a project for the analysis of emotion concepts. With Mun, I accept that the fundamental base can be initially characterised through ordinary intuitions about emotions. However, in contrast with Mun, I claim that ordinary intuitions should not include first-person experiences. Instead, I propose a pragmatic account of the fundamental base that, following insights from semantic externalism, identifies ordinary intuitions with the public use of emotion concepts. The upshot of this account is an empirically tractable project in emotion research that is compatible with an interdisciplinary science, and that raises important epistemic and ethical questions towards constructing a scientific theory of emotion.

My argument will proceed as follows. First, I will rehearse some arguments in the literature in support of the idea that emotion science should be based on ordinary intuitions or some form of folk psychology. Second, I will focus on Mun's account of the fundamental base, and show the merits and limitations of characterising the fundamental base in terms of first-person intuitions. To achieve this, I present an argument based on considerations concerning theoretical and epistemological difficulties with relying on first-person

experience for the purposes of identifying the fundamental base. Third, I present my pragmatic account of the fundamental base, drawing from Haslanger's (2005, [2000] 2012, [2006] 2012) ideas on semantic externalism and social construction. Fourth, I conclude by raising some epistemic and ethical questions for the project of investigating the fundamental base and formulating scientific concepts of emotions.

1. Emotion Science and Folk Intuitions

The question of how to define emotion concepts in science can be traced back to the question of whether emotions constitute natural kinds. The motivation behind framing the question in terms of natural kinds stems from the assumption that one of the aims of scientific inquiry is to discover how objects and phenomena in the world belong together and how they differ from others. Importantly, these distinctions are presumably independent of human interests, such that they exist independently of human practices and thus are one of the objects of discovery for science (Bird and Tobin 2022). Framed in this way, the problem is, first, to determine what explains whether objects in the world belong in a class or not—that is, to establish some criterion for natural kindhood—and second, to evaluate whether this criterion applies in the case of emotions.

There have been two important challenges to the project of identifying emotions with some natural kind (or set of natural kinds). First, there is Paul Griffiths' (1997) claim that the vernacular concept of "emotions" does not refer to any single natural kind because the phenomena captured by this concept belong to three different classes: basic emotions, arguably identified with affect programs; higher cognitive emotions, identified with evolved solutions to cooperation problems; and socially constructed emotions. Second, there is Lisa Barrett's (2006; 2017) claim that emotions do not have a neural or physiological *fingerprint*, a pattern to which each emotion category corresponds, and that they therefore cannot be said to form natural kinds or to support typological thinking (Barrett and Lida 2025). Other scientists, along with Barrett, have also published evidence that arguably casts doubt on the natural kind assumption (e.g. Lindquist et al. 2012; Touroutoglou et al. 2014).

Faced with these challenges, researchers have opted for two roads. One, defended by Griffiths, is eliminativism, i.e., the claim that there should be no "emotion science" proper, and that the vernacular concept of emotions plays no role in scientific inquiry. James Russell (2009) has defended a similar view, arguing that folk emotion concepts are not useful tools for scientific inquiry. The second kind of challenge is some form of revisionism, according to which there can be a science of emotions, but we need a separate conceptual framework to capture what it is that emotion science is studying. This new conceptual framework would presumably satisfy the demands of scientific inquiry, such as supporting explanations and predictions, and, crucially, should be allowed to differ from folk understandings of emotions. This is the idea behind Andrea Scarantino's (2012) distinction between the Folk Emotion Project and the Scientific Emotion Project.

Current emotion science leans towards this latter alternative, offering revised concepts of emotions that are presumably more scientifically fruitful than their vernacular counterparts. However, as with any analytic endeavour, any concept that is a candidate to be the *analysans* of a vernacular concept faces the risk of differing so much from it that it ends up being an entirely different concept altogether. One such objection was raised by P. F. Strawson (1963) against Rudolf Carnap's ([1950] 1963) idea that we could explicate folk concepts to construct more scientifically fruitful ones. This is also a common theme in discussions on conceptual engineering, a project that inherits from Carnap in calling for the construction of concepts that are more suitable for specific epistemic, ethical, or political projects than the currently operating (often folk) concepts (Díaz-León 2020; Thomasson 2020).

The solution to this problem is to somehow *anchor* scientific concepts in folk concepts, such that we can reasonably claim both that the new scientific concepts are better suited for empirical research and that they still refer to some extent to the phenomena we refer to when we use folk concepts. This is important to guaranteeing that emotion science remains a science of *emotion*, which is what the theory was supposed to explain in the first place. How this anchoring relation works remains to be worked out.

The problem of how to explicate the anchoring relation can be divided into two parts. First is the question of what the properties of this relation are, and which formal and epistemic criteria this relation should respect. The second is which domain such a relation is to be defined over, that is, how to characterise the domain of folk emotion concepts that will constitute the basis for scientific theory construction. I will discuss the second question throughout the rest of this article. Yet, something should also be said about the first.

In my view, the anchoring relation should first be such that it preserves important elements of the extension and the anti-extension of the original concept. In other words, scientific concepts should be similar enough in extension to folk concepts. I shall call this the *similarity criterion*, honouring Carnap's similarity criterion of explication. It is important to clarify, though, that the conditions under which an element may be included or excluded from the extension of a given category should be determined in actual scientific practice, based on theoretical values such as coherence, explanatory power, or other sets of epistemic (and even non-epistemic) criteria, rather than being decided *a priori*. This entails that deciding on limit cases and constructing scientific concepts anchored in folk concepts is an ongoing endeavour, rather than a one-off conceptual analysis project.

Second, the domain of the anchoring relation, i.e., folk concepts, must provide the grounds on which evidence will be obtained to evaluate different scientific theories of emotion. Otherwise, we would be identifying the *explananda* by presupposing theoretical constructs that already fit the explanation we want to obtain. Therefore, folk categories must be identified in ways that are as theoretically neutral as possible, at least regarding scientific theories of emotion, which will be evaluated based on evidence obtained by using these categories. I shall call this the *neutrality criterion*.

Lastly, researchers must be able to eventually agree to some extent on the reference of the folk concepts in question. This entails that it must be possible to settle disagreements that may emerge when identifying what it is that we are talking about when using folk concepts. To be able to settle such disagreements, we must have some way to ostensively pick out the phenomenon to which folk concepts refer, even if we temporarily disagree on how to describe them. This would enable us to compare and contrast our conceptions while maintaining a fixed reference, enabling us to eventually construct a unified framework in which to construct scientific concepts. I shall call this the *intersubjectivity criterion*.[1]

In sum, to anchor scientific concepts of emotion in folk concepts means to construct theoretical concepts that are similar enough in extension to folk concepts, which requires a method of identifying the latter. To identify folk concepts, we need some framework that is theoretically neutral (relative to the theories of emotion that we will evaluate based on how they respect the extensions of folk concepts) and that allows us to ostensively pick out the phenomena that emotion science is supposed to explain.

1 This criterion can also be called an *objectivity criterion* if we understand objectivity as a form of intersubjective agreement. In this sense, the intersubjectivity criterion would warrant what Douglas (2004) calls *concordant objectivity*.

With these criteria in mind, how should we identify folk concepts to which emotion science should be anchored? In the next section, I will discuss Mun's account of what she calls the *fundamental base* of emotion science. I will argue that, by relying on first-person experiences, Mun's account does not satisfy the neutrality or intersubjectivity criteria. If this is correct, a new account of the fundamental base of emotion science is required.

2. The Problem of the Fundamental Base

2.1 Mun's Account of the Fundamental Base

The claim that emotion science should be based on ordinary intuitions is what Mun (2021) calls the *fundamental base* of emotion science. In her words, it is the claim that "an adequate theory of emotion must recognise that *ordinary folk intuitions* serve the fundamental purpose of identifying the explananda for the science of emotion" (93; my emphasis).

Although Mun offers some clues as to what she means by ordinary folk intuitions, they are still difficult to define precisely. First, Mun clarifies that when she speaks of ordinary folk intuitions, she means "the *ordinary experiences* that many people have which allow them to determine not only whether or not someone is experiencing an emotion, but also (for many) which allow them to determine what kind of emotion they are experiencing to a certain extent" (94; my emphasis).

This suggests that ordinary folk intuitions are the *experiences* which allow people to identify emotional experiences in others as well as themselves. Importantly, this use of the term "intuition" differs from other standard uses in the literature which understand intuitions as beliefs, dispositions, and seemings, among others (Pust 2024).

Immediately afterwards, Mun writes: "These ordinary intuitions are an object of study for the science of emotion, and such research is typically carried out as studies on emotion recognition and attribution" (94). By emotion recognition and attribution, Mun means recognising an emotion in another and identifying an emotion type, respectively, based on facial expressions, vocal cues, and bodily gestures (I shall refer to the collection of these three cues as the "emotional expression"). To illustrate this research, Mun reconstructs the findings from DiGirolamo and Russell (2017) that show (in her interpretation) that, although forced choice paradigms in emotion recognition and attribution studies artificially inflate agreement between subjects when they are asked to categorise an emotional expression, there is still some degree of uniformity in how subjects recognise and categorise emotions in others when using open choice paradigms. For Mun, this suggests that there is some uniformity to the "intuitions" that subjects use to categorise emotional expressions, which supports the idea that emotion science should begin from such intuitions.

Finally, Mun discusses whether emotion recognition and attribution require knowing the word associated with a given emotional expression, which she denies. In her view, emotion recognition and attribution are processes that need not involve labelling an expression, but only require reactions to others' expressions and behaviour. These processes, when they occur without labelling, exemplify, in Mun's words, "[our] recognitions of our ordinary intuitions about emotions, and it is with these intuitions, along with our first-person emotional experiences, that the science of emotion must begin" (2021, 98).

At this point, we can identify some of the problems with Mun's presentation of ordinary intuitions as the basis of emotion science. As I understand Mun's view, her claim is that emotion science should begin by investigating how people recognise that others are experiencing an emotion and how they identify which emotion they are experiencing. Yet, it is unclear which role experiences—especially *first-person emotional experiences*, which she emphasises—plays in identifying the *explananda* of emotion science. The studies on emotion recognition and attribution which Mun considers, although interesting in their own right, do not offer much information about the first-person experiences which enable subjects to attribute and recognise emotions. On the face of it, these studies only offer information about the general pattern of behaviour that can be observed by a third-person observation method across many subjects, not about what their experience is or how these subjects are using their experiences to identify emotions in others. Furthermore, while ordinary intuitions were initially identified with "the ordinary experiences [which allow emotion recognition and attribution]," in the latter formulation, "ordinary intuitions" and "first-person emotional experiences" appear as separate (and the invitation is to study one "along with" the other). Hence, it is unclear whether ordinary intuitions are the same as first-personal experiences, what the role of emotion recognition and attribution studies ought to be, or how ordinary intuitions and first-personal experiences should be understood in general.

In my view, there are at least two interpretations of Mun's idea. If we take the first formulation at face value, ordinary intuitions are first-person experiences which people use to recognise and attribute emotions in others. If we take the latter presentation more seriously, ordinary intuitions are separate from first-person experiences, raising the question of what these ordinary intuitions are and how they differ from first-person experiences. Put succinctly, either ordinary folk intuitions are identified with first-person experiences or they are separate and left undefined.

Despite these obscurities, there are some ideas that can be extracted more clearly from Mun's presentation. Under either interpretation of what ordinary intuitions are, first-person experiences are a necessary part of the fundamental base of emotion science, either because they are identical to ordinary intuitions, or because they are part of the fundamental base of emotion science, "along with" ordinary intuitions (whatever they may be). Additionally, for Mun, ordinary intuitions and first-person experiences (understood either as identical or as separate but equally important) can be studied by examining emotion attribution and recognition. This implies that emotion attribution and recognition studies offer information about ordinary intuition and first-person experiences as they influence attribution and recognition.

In what follows, I will argue that the first interpretation leads to an untenable account of ordinary intuitions that does not help us to clarify the fundamental base of emotion research. This would leave us with the second interpretation, that "ordinary intuitions" are left unanalysed. If we accept this second interpretation, then the remaining task is to offer an analysis of ordinary intuitions that makes them scientifically tractable and suitable for investigation, from which to begin theorising about emotions, a task that I undertake later.[2]

2.2 Against First-Person Approaches to the Fundamental Base

Let us assume that Mun does rely on first-person experiences as necessary elements of ordinary intuitions. Does this account of the fundamental base offer an account of folk categories that satisfies the similarity, neutrality, and intersubjectivity criteria mentioned above? In what follows, I will argue that even if this account satisfies the similarity and neutrality criteria, it does not satisfy the intersubjectivity criterion. This is because what

2 This means that if my interpretation of Mun is incorrect, and what she has in mind is not that ordinary intuitions involve first-person experiences, the remainder of the argument can still be understood as an expansion on her overall project.

Mun identifies with the fundamental base, i.e., "ordinary intuitions and first-person experiences," does not provide grounds for picking out *explananda* intersubjectively. Consequently, we need some other account of the fundamental base that can satisfy all of the criteria and that offers a scientifically tractable base on which to anchor emotion science.

Let us consider how we can identify the phenomena that lie at the fundamental base according to Mun's account. Recall that Mun's fundamental base consists of the ordinary intuitions and first-person experiences that enable emotion recognition and attribution. To identify them, according to Mun's story, we can rely on the methods used in emotion recognition and attribution studies, such as forced or open choice paradigms and the like. Yet, as I argued in the previous section, what are important to (or at least a necessary part of) the fundamental base are the first-person experiences which subjects use to recognise and attribute emotions in others. The methods used in emotion recognition and attribution studies, which lead to agreement scores rather than self-reports, must be then interpreted as offering indirect information on first-person experience and how it is (consciously or unconsciously) deployed in emotion recognition and attribution.

Does first-person experience help to identify the *explananda* of emotion science? In my view, it is not possible to satisfy both the neutrality criterion and the intersubjectivity criterion mentioned before if we identify the *explananda* of emotion science with, or by means of, first-person experience. Either we need a purely phenomenal description of what occurs in emotion recognition and attribution, and first-person experience (which cannot be accessed by others, violating the intersubjectivity criterion), or we need a theory of emotion that describes our phenomenology in some way as to explain how first-person experience plays a role in emotion recognition and attribution (which violates the neutrality criterion). Hence, the appeal to first-person experience is a non-starter when it comes to identifying the fundamental base for emotion science. Let us unpack this claim.

First, let us assume that what help to identify the *explananda* of emotion science are the purely phenomenal aspects of first-person experience. To get at them, we must obtain a description of this phenomenology such that it is clear what emotion science is purported to explain. However, how can such descriptions be intersubjectively compared and contrasted so as to identify the *explananda* of a scientific research programme? Given that we do not directly share epistemic access to the first-person experience of others, this means that we would not have epistemic access to what emotion science is supposed to explain. This would violate the intersubjectivity criterion, since it would make public identification of the *explananda* of emotion science impossible (insofar as science involves not only first-person descriptions of mental states but also communities correcting and jointly investigating a common phenomenon of interest).

There seems, however, to be a way out. Even if we do not have first-person epistemic access to others' experiences, it might not be the case that first-person experiences are completely epistemically inaccessible. After all, we have third-person methods of observation that may inform us about others' experiences. Examples of these methods may include precisely the methods Mun has in mind, that is, choice methods involved in emotion recognition and attribution, among others. Interpreted as offering indirect information on first-person experiences, these methods would provide evidence of the first-person experiences of subjects, even beyond the information that can be obtained by self-report.

Yet, to make sense of how these methods give us information about first-person experiences, we already need to presuppose a theoretical background that makes explanations bridging third-person observation with first-person experience intelligible. Not only would this require, for instance, a theory of consciousness (which

Mun discusses), but also a theory of emotion that already explains how our methods of observation map onto first-person experience, and how first-person experience affects the way in which we recognise and categorise emotions in others. Hence, even if we had a workable theory of first-person experience in a form that allowed intersubjective validation (that is, which connected third-person methods of observation with first-person emotional experiences), we would be violating the neutrality criterion, i.e., we would presuppose a theoretical framework, which is what we are supposed to evaluate after (and not before) identifying the fundamental base.[3]

In sum, identifying the fundamental base of emotion research, what identifies the *explananda* of emotion science, by appealing to first-person experience, is untenable. At best, it introduces important theoretical problems, and at worst it is epistemically unsound, given the risk of circularity (i.e., of violating the neutrality criterion) or lack of intersubjective access to the phenomena under investigation (i.e., violating the intersubjectivity criterion). We need, therefore, another way to fix what would constitute the fundamental base of emotion science. In other words, we need a different way to analyse the so-called "ordinary intuitions" involved in starting, and more importantly, evaluating a scientifically tractable theory of emotion.

In what follows, I will propose one such analysis of "ordinary intuitions" that respects the neutrality and intersubjectivity criteria. What we need, I shall argue, is an ostensive device that can help make reference to the *explananda* of emotion science without presupposing a specific theory of emotions (and hence no specific explanation of the role of first-person experience in emotion recognition and attribution), while allowing for intersubjective access to the phenomenon of interest, and that can enable means of comparison that are key to any scientific enterprise. These devices, I argue, are the concepts that operate in sociolinguistic exchanges about emotions.

3. A Pragmatic Account of the Fundamental Base

In this section, I will propose an account of "ordinary intuitions" in terms of sociolinguistic practices. My claim is that so-called "ordinary intuitions" about emotions should be identified through the referents of emotion vocabularies (folk emotion concepts, descriptions, etc.) and how they operate in social practices. To unpack this claim, I will follow Sally Haslanger's (2005, [2000] 2012, [2006] 2012) semantic externalism and her analyses of folk concepts of gender and race. While I do not want to commit to the idea that emotions are social kinds,[4] I believe that Haslanger's analyses of these concepts offers a framework that works well for other types of folk concepts. This helps maintain the similarity criterion, along with the neutrality and intersubjectivity criterion, which would lead to a better account of the so-called "fundamental base" of emotion science.

3 Later, I will argue for an approach based on semantic externalism. While this may also be a theoretical framework, it does not constitute a theory of emotion, and it is plausibly less committal than a full-blown theory of phenomenal consciousness.

4 There is a fundamental difference between referring to a phenomenon in the context of social interaction and holding a phenomenon to be a social kind. Kindhood is a relation concerning what warrants similarity judgments between a group of objects, often identified with essences (Kripke 1980) or mechanisms underlying the properties these objects exhibit (Boyd 1991). Claiming that some phenomena (such as emotions) form a social kind implies the claim that these phenomena are united (epistemically or metaphysically) by social properties. Yet, referring to a phenomenon in the context of social interaction means identifying devices that refer to some objects under some description in social interaction, without specifying which properties warrant similarity between the objects. For instance, in social interactions we can refer to eyes and describe them in a myriad of ways (beautiful, aggressive, passionate, etc.), but this does not imply that they form social kinds. My claim that we can attend to how we refer to emotions in social interactions, without committing to the view that emotions form social kinds, is analogous to claims about eyes in this regard.

3.1 Haslanger on Folk Concepts

Haslanger is interested in the question of how folk concepts of social kinds relate to more technical or theoretical concepts that can be used in the social sciences and in political discourse. She studies cases such as the concept of "woman," "man," "Black," "White," and other social kind concepts where everyday use may not map onto theoretical uses of these terms (similar to the case of folk and scientific emotion concepts). For instance, when people use concepts of race in everyday discourse, they might assume that race refers to some biological grouping of human beings that, in reality, does not exist. While people who use race terms in this way might fail to refer to what they think they are referring to, it is still problematic to claim that race does not exist at all. After all, people are affected by racial categories, and therefore there appears to be a social practice that is configured around these concepts. To resist such practices, Haslanger claims that we should be sensitive to differences between what people have in their minds when they use race and other social kind concepts, and the social practices to which these concepts actually refer.

One way to understand the differences between what people think they mean and what they actually mean is by invoking a distinction between *manifest* and *operative* concepts. According to Haslanger, manifest concepts are those that are accessible to people through introspection, that is, they are what people have in their minds and can offer as a definition when asked about what they mean. Operative concepts, on the other hand, are concepts that describe actual social meanings around which people coordinate their behaviour. Importantly, manifest and operative concepts often do not coincide. This is because individuals may not be aware that what they think by the use of a given concept does not adequately capture how they are actually behaving.

Haslanger uses the illustrative example of the concept "parent." When used in the context of her daughter's school, many people have in mind a concept of "parent" as that of "progenitor." Taken at face value, this use of the concept excludes parents who did not give birth to their children, such as adoptive parents or other caregivers. However, Haslanger notes that the concept "parent," in actual social practice, does include these other forms of parenthood, since it is expected that all caregivers, regardless of their biological relation to their children, respond to calls addressed to "parents" (e.g., when invited to "Parents' Night" at school). Hence, we can distinguish between what people might have in mind when offering a definition of "parent" (i.e., the often biology-laden concept), and the concept that adequately captures the actual social practice around which the community is organising their behaviour. The first is what Haslanger calls the *manifest* concept; the latter is the *operative* concept.

How can we cash out the difference between *manifest* and *operative* concepts? Haslanger explains this difference in terms of the distinction between semantic internalism and externalism. According to semantic internalism, the meaning of a term is bound to what speakers have in mind when they use the term and what is accessible to them when explaining their meant use of the term. Hence, to study the meaning of a term, according to the internalist, one should examine through introspection what one means (e.g., speaker intentions or conceptions) when one uses a term. Semantic externalism, in contrast, is the idea that the meaning of a term is bound to external facts about the use of the term that may not be known by speakers. As Wikforss (2008) explains it, externalists claim that it is not psychological states that determine meaning, but external (physical or social) facts about the environment which may or may not be known to speakers.[5] In view of the distinction between semantic internalism and externalism, the difference between manifest and operative concepts can

5 To be clear, this is what Wikforss (2008) calls *foundational externalism*. She distinguishes two other forms of externalism, namely, *externalist semantics* and *psychological externalism*. Here I restrict externalism to *foundational externalism*, a claim about meaning determination, rather than a claim about the semantic value of terms (*externalist semantics*) or the content of mental states (*psychological externalism*).

be cashed out in terms of what determines meaning: while the meanings of manifest concepts are determined internally (by the psychological states of speakers when using a term), the meanings of operative concepts are determined externally (by physical or social facts about their environments and the linguistic communities who use them).

What kinds of concepts are amenable to analysis under the distinction between manifest and operative concepts? Haslanger thinks that to analyse social kinds, we must adopt an externalist view of the semantics of social kind terms. This is because, as illustrated by the previous example, people might not be sensitive to how a concept is actually operating in practice, and they might err about the role a concept plays when deployed within a community. In other words, their psychological states might not reflect the actual ways in which they are organising their behaviour (e.g., what they actually expect from others) when using some term, such that they might think they mean something when they expect different reactions from others (e.g. a school principal who is still including non-progenitor caregivers under the extension of "parent" and expecting certain forms of behaviour despite their explicit beliefs about the meaning of the term).

The distinction between manifest and operative concepts need not be limited to terms referring to social kinds such as "parent," even though those are Haslanger's focus. Any time our mental states about referents (i.e., intensions) may differ from what these referents (i.e., extensions) are, a distinction can be drawn between what we believe, our intentions, and what we are talking about. These are the traditional cases for semantic externalism about natural kinds such as water or arthritis (Putnam 1975; Burge 1979). Consequently, put in terms of folk theories and concepts, whenever we have a folk theory which involves some form of reference to a phenomenon (e.g., some vocabulary), we may analyse it in terms of what users of such a theory think and how such a theory operates, and we must be open to the possibility that these two levels of analysis do not match.

With Haslanger's general framework for analysing social kind concepts in mind, let us now apply it to the case of the fundamental base for the science of emotion. First, I will argue that the ordinary intuitions and folk emotion concepts relevant to approaching the fundamental base must be coined in terms of operative rather than manifest concepts. Here I will mostly follow Haslanger's strategy concerning social kinds, adding some nuances more specific to the case of emotion concepts. Second, I will present some interesting sources of evidence to which researchers can turn to examine the fundamental base. These include studies in emotion attribution and recognition, following Mun, but also other sources such as cross-cultural research on variations in emotion terms, and research on emotion scripts.

3.2 Back to the Fundamental Base

Let us take stock of the argument so far. I have argued that scientific emotion concepts must be somehow anchored to folk emotion concepts, and that this anchoring relation must satisfy three criteria: scientific emotion concepts must be similar enough to their folk counterparts (*similarity*); folk emotion concepts must be understood in theory-neutral terms relative to a theory of emotion (*neutrality*); and some form of intersubjective contrast must be possible, to settle disagreements about how we understand the fundamental base (*intersubjectivity*). I have also argued that Mun's appeal to first-person experience does not satisfy the neutrality and intersubjectivity criteria, leaving the question of how to identify the *explananda* of emotion science (i.e., the fundamental base) undetermined. How then should we identify the *explananda* of emotion science?

Let us begin with what, I believe, is a reasonable assumption, and one in the spirit of Mun's idea of the fundamental base: emotion science should begin by identifying the phenomena we usually make reference to by using emotion vocabulary (the term "emotion," emotion categories, descriptions, etc.). One traditional approach to such questions was to investigate these terms and their respective concepts *a priori*, but this would, at best, reveal beliefs about emotions rather than clarify what we are talking about (i.e., some subset of manifest concepts of emotion). Following the insights that follow from semantic externalism, and Haslanger's ideas concerning manifest and operative concepts, I propose we identify the *explananda* of emotion science with the referents of *operative concepts* of emotion.[6] In other words, investigating the fundamental base of emotion science is to investigate the phenomena to which communities refer with their use of vocabulary such as "angry," "sad," "happy," and so on, which involves understanding how such vocabulary operates in practice (rather than what we believe about these terms).[7]

Investigating operative concepts, as with other projects of discovering what we are talking about, is an empirical matter. We must investigate when these concepts are applied, consciously or not, and to which sorts of phenomena they make reference. Crucially, rather than taking them at face value, these concepts must work as *ostensive devices*, as means by which we signal what we mean. To clarify this, consider the following analogy: if we want to investigate water, we should attend to the substance referred to by the term "water," using the term as an ostensive device to make reference to (rather than describe) the phenomenon that we are interested in. We would not take beliefs about water at face value, but as ways of making the reference explicit so that we can be sure what it is that we are supposed to investigate.

Appealing to semantic externalism in this way allows the investigation of the fundamental base to satisfy the criteria laid down above. First, we are guaranteed to maintain some similarity between folk and scientific concepts, insofar as reference must be maintained in both domains. In other words, by maintaining referents fixed for the clearer cases under the extension of folk concepts, scientific concepts will keep sufficient similarity with their folk counterparts. Second, by taking folk concepts as ostensive devices, and not as informing us about the nature of emotions themselves, we are respecting the neutrality criterion, since we are not presupposing a theoretical framework to understand emotions, nor one that explains the role some other phenomenon (such as first-person experience) plays in emotional processes. Lastly, given that we are fixing reference in terms of the public application of emotional vocabulary, the account I'm proposing satisfies the intersubjectivity criterion by providing means of reference that are publicly accessible and that allow disagreement and consensus on what is it that emotion science should explain.

To make this proposal clearer, I will dedicate the remainder of this article to illustrating how research on the fundamental base can advance by considering different aspects of how emotion concepts, vocabularies, terms, descriptions, and the like, operate. I will focus on three sources of information about how these vocabularies operate in practice, namely: emotion recognition and attribution studies (reconsidered), cross-cultural variations of emotion vocabularies, and research on emotion norms and scripts. This is by no means an exhaustive list of potential sources of evidence, but an illustration of how to expand the examination of the fundamental base to other areas of research and disciplines.

6 Another alternative is to assume that we can identify reference by identifying natural kinds of emotions (see e.g. Scarantino 2012). Yet, this would already assume that emotions form natural kinds, an assumption that is contentious in the literature (Barrett 2006). While I believe some version of natural kind externalism about emotions is correct, I will not expand on it here, and I will assume a more modest approach to what unifies emotion categories in science.

7 This vocabulary need not only involve names for emotion categories, but may also include descriptions and other linguistic forms of reference.

3.3 Emotion Attribution and Recognition

Let us begin by discussing the two sources already mentioned by Mun, namely, studies on emotion attribution and emotion recognition. I have argued that, on Mun's account, emotion attribution and recognition studies must be interpreted as providing information about first-person experience. Given the externalist framework I am proposing, which rejects first-person experience as fundamental to identifying the *explananda* of emotion science, should emotion recognition and attribution studies still be considered part of the investigation of the fundamental base?

One way of using emotion attribution and recognition studies is to interpret them, not in terms of how they inform us about first-person experience, but in terms of what they inform us concerning operative concepts of emotion. Following DiGirolamo and Russell (2017), for instance, we can claim that folk emotion categories are relatively stable despite subjects choosing among open-ended options. This does not imply that emotions are, by nature, stable kinds, but rather that the *explananda* of emotion science involve some degree of stability. This approach mirrors what Scarantino (2012) suggests as the Folk Emotion Project, which interprets findings in experimental psychology as findings concerning folk concepts of emotion. Making such interpretations clear is important, since appealing to emotion attribution and recognition studies would otherwise presuppose a set of theoretical concepts which would violate the neutrality criterion.

Other findings in emotion attribution and recognition studies may be interpreted along similar lines. For example, empirical evidence suggests that both emotion attribution and recognition are highly context-sensitive. Sabini and Silver (2005) show that distinctions between shame and embarrassment, as well as between envy and anger, are sensitive to contextual factors. According to the researchers, subjects' judgements of shame and embarrassment are differentiated depending on whether they perceive that an event has revealed a real flaw in themselves (in the case of shame) or has only appeared to reveal a flaw, but the flaw is not considered real (in the case of embarrassment). In the cases of envy and anger, envy is attributed in cases where a person unwarrantedly accuses another person's accomplishment, while anger is attributed in cases where such an accusation is taken as warranted. In both cases, there are social norms that dictate whether a flaw is considered real or not, or whether an accusation is considered warranted or not, which in turn determine the emotion to be attributed. Similar studies include findings that show that subjects use contextual information even under the instruction to disregard context (Ngo and Isaacowitz 2015), that subjects suffering from depression are much less sensitive to contextual cues, which affects their accuracy in emotion recognition (Rottenberg et al. 2005), and that subjects are prone to using contextual information to disambiguate between possible emotions to attribute in a given situation (Hareli et al. 2018).

Overall, the case for context-sensitivity effects seems well-supported by empirical evidence in experimental psychology. Interpreted as offering information about the operation of folk concepts, these findings shed light on how emotion concepts operate in actual practices and help us to identify the phenomena that emotion science is supposed to explain and theorise about. Yet, these are not the only studies that help us understand the operation of folk emotion concepts or that influence our understanding of the fundamental base of emotion science.

3.4 Emotions and Cross-Cultural Linguistics

Another important source of empirical evidence relevant to investigating the fundamental base are studies on how different emotion terms operate across cultures. If part of the fundamental base concerns the tools with which people describe their emotions, investigating the use of emotion terms seems quite straightforward.

Yet, one important aspect to emphasise is that such an investigation must be carried out relative to specific languages. This is because there are considerable difficulties in translating emotion terms across languages, and while these difficulties should be addressed, differences between linguistic contexts must be acknowledged. For example, a considerable amount of emotion research has presupposed that folk concepts coming from English are the basic starting point for a science of emotion. This yields a limited research programme that does not properly account for differences across languages, which at the very least leads to dire epistemological issues for theorising about emotions. This line of criticism has been pressed with great emphasis by Anna Wierzbicka (1999; 2009; 2014; see also Levisen 2019).

Wierzbicka has shown several cases where translation of emotion terms across cultures fails. For example, the English term "sadness" translates into two terms in Russian, namely, "grust" and "pečal." Further research on translation has shown similar results. Barger et al. (2010) show that translating "disgust" into Chinese yields four different terms: "taoyan," "yanwu," "exin," and "otu." However, when back-translating these terms into English, the researchers note that only "exin" translates to "disgust," although it contains themes related to "anger" as well, and "yanwu" and "taoyan" seem to be inadequate translations, missing some core themes of the English term "disgust." This suggests that English terms do not map neatly onto terms in other languages.

Research on the fundamental base of emotion science should take such translation problems seriously. This is because, if we prioritise English vocabularies, "sadness" is bound to be considered a single discrete category; but if we prioritise Russian vocabularies, we would be motivated to split this category into two types of emotion. Such decisions inform further taxonomical and empirical questions and are therefore a vital part of the investigation of the fundamental base.

Wierzbicka's solution is to construct a minimalistic metalanguage that enables translation using universal semantic primes, a metalanguage that she calls the *Natural Semantic Metalanguage* (NSM). As Goddard describes it: "Using the NSM metalanguage allows us to decompose complex language-specific concepts into configurations of simple concepts that are shared across languages. This allows a very high degree of semantic resolution and enables us to access language-specific meanings using rigorous, evidence-based procedures of semantic analysis" (2015, 294).

By investigating how emotion concepts vary across cultures, we can understand how wide the fundamental base of emotion science should be. Given Wierzbicka's findings, we should work towards a science of emotion that is not limited to how English emotion concepts operate. This exemplifies how research in cross-cultural linguistics can inform what we consider to be at the base of the science of emotion.

3.5 Emotion Scripts and Norms

Cross-cultural variations in emotions are not limited to variations in emotion terms. Emotional behaviours also vary across cultures. These include variations in expression, action tendencies, and social norms attached to emotional responses. These variations can be summed up under a common concept of variations in emotional *scripts* (Wierzbicka 1994; Eickers and Prinz 2021; Eickers 2024).

Eickers defines scripts as follows: "Scripts are normative, context-sensitive, nested knowledge structures that describe behavior in terms of corresponding events, situations, social roles, individuals, or mental state types in a way that guides action" (2024, 7–8). These structures are paramount to social cognition, as they are determinants of how people behave regarding social norms. In the case of emotions, scripts involve our

knowledge of how emotional expressions, action tendencies, and other aspects of emotional behaviour, occur in concrete social settings. Hence, investigating emotional scripts can offer important information on operative emotion concepts that constitute the fundamental base, since they are part of the practices to which such concepts refer.

There are two aspects of scripts that are worth mentioning for the present argument. First, scripts do not necessarily refer to internal representational states or beliefs social agents have about the social world. This is important because it marks the difference between using them to investigate operative concepts of emotion and using them to investigate manifest concepts. While scripts can be made explicit and used consciously by individuals (e.g. when I learn how to behave in a specific context by learning the appropriate sequence of acceptable behaviours), they need not be used consciously. Most of our social lives actually occur without us reflecting on our behaviour, and emotional behaviours are no exception. This supports the externalist view I am arguing for, since, by investigating scripts, we can make sense of factors pertaining to emotional behaviour beyond beliefs and other internal states of individual minds.

Second, emotion scripts support and expand on both of the aforementioned areas of research, namely, investigating forms of emotion attribution and recognition from a cross-cultural perspective. Scripts are highly context-sensitive and culturally varied, and they help explain variations in the context of emotions. As Eickers and Prinz put it:

> Evidence for learning of scripts can be found in the aforementioned fact that emotions differ cross-culturally. For example, there are culturally specific behaviors such as bowing, flicking-off, and clapping. In addition, cultures differ in whether they encourage people to act aggressively when angry or to quietly sulk or brood (see Goddard (1996) on Malaysia), while other cultures discourage brooding (see Briggs (1970) on the Ifaluk). This suggests that anger is not an automatic response with a fixed action tendency, but rather a role that we act out in culturally prescribed ways. (2021, 356)

Besides the evidence Eickers and Prinz mention, there are other studies that exemplify how emotion scripts can serve to examine folk emotion concepts. For instance, Uchida and Kitayama (2009) show that US American and Japanese descriptions of happiness differ in how positive these emotions are and how much they are related to social relations. For US American populations, happiness is much more related to personal achievements and positive hedonic experiences than for Japanese populations, who judge happiness to involve ambivalent experiences that can even be socially disruptive.

Not only do cultural variations in emotion scripts help to account for variations in emotion terms, but they are also an important piece of emotion recognition and attribution. As I explained above, emotion recognition and attribution are context-sensitive. One way to understand such effects is by appealing to emotional scripts. If emotional scripts are part of the knowledge structure deployed to recognise and attribute emotions to others, and such structures involve the use of contextual information to specify and disambiguate information in the behaviour of others, it is natural to conclude that context-sensitivity in emotional attribution and expression is due to the way these scripts operate when people are engaging in these processes. In Eickers and Prinz's terms, emotion recognition is, after all, a social skill.

In sum, I take emotion recognition and attribution, emotion terms, and emotion scripts to be constitutive of the fundamental base of emotion research, the so-called "folk intuitions" at play that underlie emotion

science and that are intended to anchor theoretical concepts of emotions. This invites an investigation of the fundamental base that is interdisciplinary, follows Mun's account of the relation between folk emotion concepts and scientific theories, and also expands to add other areas of research that can contribute to our understanding of emotions.

4. Conclusion

In this paper, I have argued for an account of the nature of the fundamental base of emotion science and a way to examine folk intuitions about emotions. While Mun (2021) advocates for grounding emotion science in ordinary intuitions, there are important philosophical challenges in determining the nature of these intuitions. Specifically, there are reasons to exclude first-person experiences from the scope of the fundamental base of emotion science. Instead, I posit an alternative approach that expands on Mun's, arguing that the fundamental base for emotion science should be characterised by the public use of emotion concepts. This pragmatic stance aligns with the goal of establishing an empirically tractable framework in which to theorise about emotions.

I started by first presenting existing viewpoints supporting the reliance on ordinary intuitions or folk psychology in emotion science. Subsequently, I examined Mun's account of the fundamental base, highlighting both its merits and limitations, particularly concerning first-person experiences. I focused on concerns related to how relying on first-person experience to identify the fundamental base does not satisfy the neutrality and intersubjectivity criteria that an account of the *explananda* of a scientific research programme should satisfy. These arguments led to a pragmatic account informed by Haslanger's ideas on semantic externalism and social construction, and which, in my view, offers a promising alternative that circumvents the challenges to the use of first-person experiences in identifying the *explananda* of emotion research.

On this pragmatic account, the explananda of emotion science should be identified by investigating operative concepts of emotion—that is, how folk emotion concepts operate in actual sociolinguistic practices. This approach shifts the focus from individual subjective experiences to shared, public understandings of emotions, which can provide a more stable foundation for emotion research. By grounding emotion science in these operative concepts, we can avoid the pitfalls of relying on first-person experiences, which are inherently subjective and difficult to validate across different individuals.

Furthermore, this pragmatic stance not only addresses methodological challenges, but also has broader implications for the interdisciplinary nature of emotion research. I exemplified how research on folk emotion concepts and the fundamental base can unfold from an externalist perspective. First, I reconsidered the role of studies on emotion attribution and recognition, suggesting that we interpret these findings as concerning the operation of folk emotion concepts rather than as indicators of first-person experiences. Second, I presented an argument from cross-cultural linguistics about how we can broaden the scope of emotion science beyond English folk emotion concepts. Lastly, I considered research on emotion scripts and norms, which provides an understanding of shared social conventions concerning emotional phenomena.

As a final note, someone might think that the account I have offered relies on an unfair misrepresentation of Mun's original idea concerning ordinary intuitions and emotional experience. While I have offered reasons to believe that, in Mun's account, ordinary intuitions and first-personal experiences are problematically intertwined, perhaps there is some other interpretation that can overcome the problems I have raised. For

instance, first-personal experiences might not be as central as I understand them to be to determining the referents of emotion words and expressions, despite textual evidence to the contrary. In that case, I would urge the reader to consider my proposal as an expansion of Mun's ideas, to include insights from semantic externalism, cross-cultural linguistics, and social cognition into the study of the fundamental base. Put succinctly, either my argument overcomes an unclarity that is present in Mun's account, or expands on a correct version of her work.

In conclusion, while Mun's work provides valuable insights into the importance of grounding emotion science in ordinary intuitions, my proposed pragmatic account offers a more comprehensive framework. By moving beyond first-person experiences and instead focusing on public, shared understandings of emotions, we can create a foundation for emotion research that is both empirically tractable and interdisciplinary. This shift not only resolves key philosophical challenges, but also opens new avenues for addressing critical epistemic and ethical questions in the field of emotion research.

5. References

Barger, B., R. Nabi, and L. Y. Hong. 2010. "Standard Back-Translation Procedures May Not Capture Proper Emotion Concepts: A Case Study of Chinese Disgust Terms." *Emotion* 10 (5): 703–711.

Barrett, L. F. 2006. "Are Emotions Natural Kinds?" *Perspectives on Psychological Science* 1 (1): 28–58.

———. 2017. "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization." *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Barrett, L. F. and T. Lida. 2025. "Constructionist Theories of Emotions in Psychology and Neuroscience." In *Emotion Theory: The Routledge Comprehensive Guide. Volume I: History, Contemporary Theories, and Key Elements*, edited by A. Scarantino, 350–87. Routledge.

Bird, A. and E. Tobin. 2022. "Natural Kinds." In *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta and U. Nodelman. Spring 2023 edition. https://plato.stanford.edu/archives/spr2023/entries/natural-kinds/.

Boyd, R. 1991. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61: 127–48.

Briggs, J. L. 1970. *Never in Anger: Portrait of an Eskimo Family*. Harvard University Press.

Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–121.

Carnap, R. (1950) 1963. *Logical Foundations of Probability*. The University of Chicago Press.

Díaz-León, E. 2020. "Descriptive vs. Ameliorative Projects: The Role of Normative Considerations." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 170–88. Oxford University Press.

DiGirolamo, M. A. and J. A. Russell. 2017. "The Emotion Seen in a Face Can Be a Methodological Artifact: The Process of Elimination Hypothesis." *Emotion* 17 (3): 538–46.

Douglas, H. 2004. "The Irreducible Complexity of Objectivity." *Synthese* 138 (3): 453–73.

Eickers, G. 2024. "Scripts and Social Cognition." *Ergo* 10 (54).

Eickers, G and J. Prinz. 2021. "Emotion Recognition as a Social Skill." In *The Routledge Handbook of Philosophy of Skill and Expertise*, edited by E. Fridland and C. Pavese, 347–61. Routledge.

Goddard, C. 1996. "The 'Social Emotions' of Malay (Bahasa Melayu)." *Ethos* 24 (3): 426–64.

———. 2015. "The Complex, Language-Specific Semantics of 'Surprise'." *Review of Cognitive Linguistics* 13 (2): 291–313.

Griffiths, P. E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. University of Chicago Press.

Hareli, S., S. Elkabetz, and U. Hess. 2018. "Drawing Inferences from Emotion Expressions: The Role of Situative Informativeness and Context." *Emotion* 19 (2): 200–208.

Haslanger, S. 2005. "What Are We Talking About? The Semantics and Politics of Social Kinds." *Hypatia* 20 (4): 10–26.

———. (2000) 2012. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" In *Resisting Reality: Social Construction and Social Critique*, edited by S. Haslanger, 221–47. Oxford University Press.

———. (2006) 2012. "What Good Are Our Intuitions? Philosophical Analysis and Social Kinds." In *Resisting Reality: Social Construction and Social Critique*, edited by S. Haslanger, 381–405. Oxford University Press.

Levisen, C. 2019. "Biases We Live By: Anglocentrism in Linguistics and Cognitive Sciences." *Language Sciences*, 76 (November 2019): 101173.

Lindquist, K. A., T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett. 2012. "The Brain Basis of Emotion: A Meta-Analytic Review." *Behavioral and Brain Sciences* 35 (03): 121–43.

Kripke, S. A. 1980. *Naming and Necessity*. Harvard University Press.

Mun, C. 2021. *Interdisciplinary Foundations for the Science of Emotion: Unification without Consilience*. Springer International.

Ngo, N. and D. M. Isaacowitz. 2015. "Use of Context in Emotion Perception: The Role of Top-Down Control, Cue Type, and Perceiver's Age." *Emotion* 15 (3): 292–302.

Putnam, H. 1975. 'The Meaning of "Meaning"'. In *Mind, Language and Reality: Philosophical Papers, Volume 2*, edited by H. Putnam 215–71. Cambridge University Press.

Pust, J. 2024. "Intuition." In *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta and U. Nodelman. Fall 2024 edition. https://plato.stanford.edu/archives/fall2024/entries/intuition/.

Rottenberg, J., J. J. Gross, and I. H. Gotlib. 2005. "Emotion Context Insensitivity in Major Depressive Disorder." *Journal of Abnormal Psychology* 114 (4): 627–39.

Russell, J. A. 2009. "Emotion, Core Affect, and Psychological Construction." *Cognition & Emotion* 23 (7): 1259–83.

Sabini, J. and M. Silver. 2005. "Why Emotion Names and Experiences Don't Neatly Pair." *Psychological Inquiry* 16 (1): 1–10..

Scarantino, A. 2012. "How to Define Emotions Scientifically." *Emotion Review* 4 (4): 358–68.

Strawson, P. F. 1963. "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, edited by P. A. Schilpp, 503–18. Open Court.

Thomasson, A. L. 2020. "A Pragmatic Method for Normative Conceptual Work." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 435–58. Oxford University Press.

Touroutoglou, A., K. A. Lindquist, B. C. Dickerson, and L. F. Barrett. 2014. "Intrinsic Connectivity in the Human Brain Does Not Reveal Networks for 'Basic' Emotions." *Social Cognitive and Affective Neuroscience* 10 (9): 1257–65.

Uchida, Y. and S. Kitayama. 2009. "Happiness and Unhappiness in East and West: Themes and Variations." *Emotion* 9 (4): 441–56.

Wierzbicka, A. 1994. "Emotion, Language, and Cultural Scripts." In *Emotion and Culture: Empirical Studies of Mutual Influence*, edited by S. Kitayama and H. R. Markus, 133–96. American Psychological Association.

———. 1999. *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge University Press.

———. 2009. "Overcoming Anglocentrism in Emotion Research." *Emotion Review* 1 (1): 21–23.

———. 2014. *Imprisoned in English: The Hazards of English as a Default Language*. Oxford University Press.

Wikforss, Å. 2008. "Semantic Externalism and Psychological Externalism." *Philosophy Compass* 3 (1): 158–81.