

Darwin on the Nature of the Emotions

In the previous three chapters concerning Wittgenstein I showed that the meanings of the terms that refer to psychological phenomena, such as emotions, desires, feelings, and sensations, are learned and taught by means of the behavior and circumstances or simply behavior that are the publicly observable defining criterion for these terms. Terms such as “jealousy” and “sadness” must be defined in terms of the behavior and circumstances or the behavior that are their defining criterion. Feeling jealousy is feeling like hurting the person with whom one had an exclusive love relationship because she has taken on a love relationship with someone else. Feeling sadness is feeling like crying and exhibiting other sadness behavior. We can be jealous or sad without being aware that we have the emotion or the feeling. Emotions and feelings are not primarily introspectable phenomena, but are responses, both internal and behavioral, to circumstances or states of one’s body. These responses need not be conscious; we may or may not be aware of our emotions and feelings. Persons who are aware that they have an emotion or feeling can often suppress almost all of the overt behavior. People can have a conscious or an unconscious emotion even when they are not in the circumstances that are part of the criterion for that emotion, e.g., when they are merely remembering or thinking about those circumstances. One can have an emotion or feeling without being aware of it and without behaving in a way that is the behavioral defining criterion for a feeling, and not being in circumstances that are the circumstantial defining criterion of the emotion. Nonetheless, the meanings of the words that refer to these emotions and feelings are determined by the observable behavior and circumstances or simply behavior that are their defining criterion.

It may seem implausible to conclude anything significant about the nature of psychological phenomena, e.g., emotions, feelings, and sensations, from the meaning of terms we

use to refer to these phenomena. However, as I showed in the previous chapters, it is impossible to understand how children can know what psychological phenomena they have without understanding how they learned the words referring to these psychological phenomena. Unlike terms that refer to physical objects and properties, the referents of psychological terms play no role in determining or teaching the meaning of psychological terms or even in determining what the terms refer to. Whereas with physical objects and properties we can show or teach a child what is being referred to by picking it up or pointing to it; we cannot show or teach a child what psychological phenomena are being referred to in this way. Rather the behavior and/or circumstances that are the defining criterion for those psychological terms must be used to show or teach what psychological phenomena are being referred to. Nonetheless it is important to provide non-linguistic arguments for regarding emotions and feelings as internal and behavioral responses to circumstances or states of the body that are sometimes introspected rather than as introspectable states that sometimes lead to behavior. I shall present some of these arguments by examining Darwin's work on the expression of emotions. Although I accept Darwin's general account of the expression of emotions, I will show how some of his descriptions of what is going on lead to a misleading account of the emotions. One reason his account of the emotions is misleading may be that Darwin holds that acquired habits of behavior can be inherited, but I believe a more important reason is his lack of concern with language.

In his book, *The Expression of the Emotions in Man and Animals*, Darwin successfully shows that there is a close relationship between the expression of some emotions in human beings and the expression of those same emotions in animals. According to Paul Ekman who edited the Definitive Edition of Darwin's *The Expression of the Emotions in Man and Animals*, (1998) Darwin "wrote this book to challenge Bell's claim that some of our facial muscles were given by God to express human emotions." He supports this claim by quoting from a letter that

Darwin wrote to Alfred Russel Wallace, “I want, anyhow, to upset Sir C. Bell’s view ... that certain muscles have been given to man solely that he may reveal to other men his feelings.” (p. 8) Darwin starts Chapter I with the following sentence, “I will begin by giving the three principles, which appear to me to account for most of the expressions and gestures involuntarily used by man and the lower animals, under the influence of various emotions and sensations.” (p. 33)

The first two principles are *The principle of serviceable associated habits* and *The principle of antithesis*. (p. 34) Darwin claims that both of these principles are related to the state of mind of the person or animal expressing that emotion. He describes the first principle as follows.

“Certain complex actions are of direct or indirect service under certain states of the mind, in order to relieve or gratify certain sensations, desires etc., and whenever the same state of mind is induced, however feebly, there is a tendency through the force of habit and association for the same movements to be performed, though they may not then be of the least use. Some actions ordinarily associated through habit with certain states of mind may be partially repressed through the will, and in such cases the muscles which are least under the separate control of the will are the most liable still to act, causing movements which we recognize as expressive. In certain other cases the checking of one habitual movement requires other slight movements and these are likewise expressive.”

In this paragraph Darwin clearly holds that the expression of emotions are the result of the actions taken under certain states of mind. So he seems to regard the emotions themselves as the states of mind that cause the actions that come to result in the expression of those emotions.

This view about the relationship of expressions of emotions to the states of mind that are the emotions is also put forward in the second principle, *The principle of antithesis*. Darwin

describes this principle as follows, “Certain states of mind lead to certain habitual actions, which are of service, as under our first principle. Now when a directly opposite state of mind is induced, there is a strong and involuntary tendency to the performance of movements of a directly opposite nature, though these are of no use; and such movements are in some cases highly expressive.” (Ibid.) In both of these principles it is clear that the expressions of emotion are the lingering results of the habitual actions that are brought about by certain states of mind. This is not true of the third principle, “*The principle of actions due to the constitution of the nervous system, independently from the first of the will, and independently to a certain extent of habit.*” Darwin says, “The third principle may, for the sake of brevity, be called that of the direct action of the nervous system.” (Ibid.)

What is evident from the first two principles is how much weight Darwin ascribes to habit. He acknowledges this force when he says, “With respect to our first principle, it is notorious how powerful is the force of habit.” (p. 35) What is remarkable about this claim about the force of habit is that Darwin applies it not only to human beings and animals such as horses, dogs, but also to pigeons, and even to moths. “To those who admit the gradual evolution of species, a most striking instance of the perfection with which the most difficult consensual movements can be transmitted is afforded by the humming-bird Sphinx-moth (*Macroglossa*): for this moth, shortly after its emergence from the cocoon, as shown by the bloom on its unruffled scales, may be seen poised stationary in the air, with its long hair-like proboscis uncurled and inserted into the minute orifices of flowers, and no one, I believe, has ever seen this moth learning to perform its difficult task, which requires such unerring aim.” (Ibid.) This passage is one of many that show that Darwin, contrary to the commonly held view of him, believed that, at least with regard to behavior, evolution involved the inheritance of habits that had been acquired by previous generations.

He explicitly says, “When there exists an inherited or instinctive tendency to the performance of an action, or an inherited taste for certain kinds of food, some degree of habit in the individual is often or generally requisite. (p. 36) Darwin not only attributes the paces of horses and the pointing of dogs to inherited habit, but even claims that inherited habit applies to the taste preferences of caterpillars. “Caterpillars which have been fed on the leaves of one kind of tree have been known to perish from hunger rather than to eat the leaves of another tree, although this afforded them their proper food, under a state of nature, and so it is in many other cases.” The view about the inheritance of acquired characteristics has generally been attributed to Lamarck, and this Lamarckian view has often been contrasted with the Darwinian view about evolution by natural selection. However, we now know that Darwin had not read Mendel’s work on genetics and so he did not know about the genetic theory of how traits were passed on from one generation to another. According to the Darwinian account of natural selection, as long as there are many different traits passed on to from one generation to another it makes little difference how these traits are passed on, by genes or by acquired habits. As long as some of these traits are better suited to the environment in which the next generation finds itself, evolution by natural selection applies to the inheritance of acquired characteristics as well as it applies to inheritance by genetic transfer.

Two other examples show how completely Darwin accepted the inheritance of habitual behavior. “It was obviously impossible that a carefully-guarded infant could have learnt by experience that a rattling sound near its eyes indicated danger to them. But such experience will have been slowly gained at a later age during a long series of generations; and from what we know of inheritance, there is nothing improbable in the transmission of a habit to the offspring at an earlier age than that at which it was first acquired by the parents.” (p. 45) In describing the behavior of kingfishers, he says, “kingfishers, when they catch a fish, always beat it until it is

killed; and in the Zoological Gardens, they always beat the raw meat, with which they are sometimes fed, before devouring it.” (p.53) He says such habitual movements “are often or generally inherited; and they then differ but little from reflex actions.” (p. 54) Darwin’s two principles are no longer acceptable because they depend on the inheritance of acquired habits. Darwin’s acceptance of the inheritance of acquired traits does not cast any doubt on his theory of evolution by natural selection; however, it does have significant consequences for the way that he accounts for the expression of emotion. Giving up Darwin’s Lamarckian account of the inheritance of acquired habits has dramatic consequences for understanding the nature of the emotions.

We now know that acquired habits are not inherited, so if we are to retain Darwin’s insight about the relationship between the expression of emotion and the natural overt behavioral responses that are the behavioral defining criterion of emotions, we must provide a different explanation than the two principles that Darwin put forward. Before modifying Darwin’s first principle, *The principle of serviceable associated habits*, it is important to notice that his first principle does not explain why or how “Certain complex actions are of direct or indirect service under certain states of the mind, in order to relieve or gratify certain sensations, desires etc.” Darwin simply takes it as given that the complex action of e.g., opening one’s eyes very widely and running away, is of direct or indirect service under a certain state of mind e.g., the emotion or feeling of fear, by relieving or gratifying certain sensations, desires, etc.

It may seem obvious that running away is related to fear in the way Darwin’s first principle states, but this relationship needs to be explained. If fear is taken as an introspected state of mind, it is not obvious why any particular action should be of direct or indirect service when one has this introspected emotion or feeling. Why does running away relieve or gratify certain sensations, desires, etc, unless these sensations and desires are closely related to

dangerous circumstances? It seems far more likely that what we now call expressions of fear and fear behavior are beneficial behavioral responses to danger independent of any introspected state of mind. Similarly, it seems far more likely that sadness expressions and behavior e.g., crying, moping, etc., are natural responses to loss. Darwin's principle of natural selection makes it far more likely that the natural behavioral sadness response is beneficial for an animal suffering a loss than that this behavior relieves or gratifies the sensations and desires related to the state of mind of sadness. Darwin simply assumes that opening one's eyes widely is of direct or indirect service under the state of mind of fear, or that moping is of direct or indirect service under the state of mind of sadness. He then mistakenly claims that the inheritance of habits explains why, fear and sadness come to have the behaviors and expressions that they do.

Darwin's first principle assumes that certain complex actions, e.g., running away or hiding, are natural responses to dangerous circumstances, and then explains why these responses give rise to what we now regard as expressions of fear. This first principle does not and cannot explain why the introspected emotion or feeling of fear should lead to the complex actions involved in fear behavior. The correct explanation of how the introspected feeling of fear, the expressions of fear and fear behavior, or the introspected feeling of sadness, the expressions of sadness, and sad behavior are related requires a complete change from Darwin's way of regarding their relationship. The relationship that Darwin puts forward in his first principle, is the standard view, viz., that the introspected emotion or feeling of fear, leads to the appropriate overt fear behavior, which because of the internal bodily changes that must precede that behavior leads to what we now regard as expressions of fear. Darwin provides the correct account of how fear behavior explains fear expressions but Darwin simply assumes the view that a state of mind explains the behavior. He assumes that view because he accepts a Lamarckian account of the

inheritance of habits; he never argues for this point and in fact it is incompatible with the genetic explanation of the inheritance of behavior.

Eliminating all mention of habits from Darwin's first principle, *The principle of serviceable associated habits*, requires eliminating all mention of states of mind, because Darwin holds that it is states of mind that caused the actions that when repeated became habits. Darwin's view about states of mind leading to actions that become habits and are then inherited is now known not to be the correct explanation of the behavior of Sphinx-moths or the gaits of horses, or of human beings. The current account of continuous genetic mutation and inheritance has replaced inherited habits that originated with states of mind as a correct explanation of the instinctual behavior of Sphinx-moths and the gaits of horses, and of human beings. According to the genetic account, mutations of genes account for changes, not only in physical characteristics, e.g., size and shape, but also in behavior. Sphinx-moths that genetically inherited mutations for behavior that were better at getting their long hair-like proboscis uncurled and inserted into the minute orifices of flowers survived and reproduced at a higher rate than those who did not inherit those mutations. Continuing beneficial mutations finally resulted in the level of behavior that Darwin commented on.

Very few mutations are beneficial; rather most are harmful. The offspring that inherit harmful mutations, i.e., mutations that result in the offspring dying before reproducing, or in being unable to produce offspring, do not pass along these mutations. Offspring that inherit mutations that are neither helpful nor harmful would reproduce at the same rate as those that inherit genetic material that has undergone no mutations. However, if at some later time it turns out that in a new environment the offspring with the new mutations survive and reproduce at a higher rate than those with no mutations, the offspring with the helpful mutations would eventually completely replace the offspring with no mutations. Or if there were contiguous

environments such that the offspring with the new mutations survived and reproduced at a higher rate in one environment and the offspring with no mutations survived and reproduced at a higher rate in the other environment, then two distinct species might evolve. The details are obviously more complex than outlined above but this oversimplified account does not distort the general picture of how evolution occurs.

Mutations are constantly occurring and though the number of helpful mutations is a very small percentage of the number of mutations, they occur with sufficient frequency that over countless generations, offspring with a helpful mutation will develop further helpful mutations. Mutations do not have to be very helpful for those with the helpful mutation to completely replace the offspring with no new mutation. These continuing small improvements, over a sufficiently long time, can and do result in astounding changes. But the time involved must be extremely long for this kind of change to take place, so the theory of evolution became plausible only after the discoveries of geologists about the extraordinary length of time that the earth has existed. Once it is granted that the earth is about four and half billion years old, and that life on earth began about a billion years later, it becomes quite plausible that small mutations can result in changes that continue over time and may even result in human beings evolving from very primitive biological forms. It also becomes plausible that some species would develop the extraordinarily complex behavior exhibited by Sphinx-moths. There is no need to talk about the inheritance of habits, and indeed with regard to Sphinx-moths such talk does not seem plausible. All that is needed is enough time for small genetic mutations to result in a continuing improvement in the behavior of Sphinx-moths.

Darwin seems to regard all actions that produce a desired goal to have been actions that were first done voluntarily in order to produce that goal. He says the following of the headless frog that used to be a staple of beginning biology courses, "It is scarcely credible that the

movements of a headless frog when it wipes off a drop of acid or other object from its thigh, and which movements are so well co-ordinated for a special purpose, were not at first performed voluntarily, being afterwards rendered easy through long continued habit so as at last to be performed unconsciously, or independently, of the cerebral hemispheres.” (p. 46) This view about the primacy of voluntary action may seem to be inconsistent with the Darwinian principle of natural selection, but it is not. Darwin says, “ It further deserves notice that reflex actions are in all probability liable to slight variations, as are all corporal structures and instincts: and any variations which were beneficial and of sufficient importance would tend to be preserved and inherited . . . although some instincts have been developed simply through long-continued and inherited habit, other highly complex ones have been developed through the preservation of pre-existing instinct – that is, through NATURAL SELECTION.” (p. 47) This passage shows that there is no incompatibility between the inheritance of acquired characteristics and natural selection.

Darwin is interested only in showing how evolution can account for the expression of emotions; he is not interested in understanding the nature of the emotions. Although Darwin was a Lamarckian with regard to emotions, and more generally with regard to inherited behavior, he understood that at the present time, the behavior is instinctual, that is, without thinking. Darwin held that what I am calling an emotional behavioral response, e.g., fleeing from a dangerous situation, was first done intentionally, but after being performed over countless generations it became habitual and then finally became instinctual. Although he was mistaken concerning the mechanism of the inheritance of behavioral responses, he realized that at the present time, most emotional behavioral responses are instinctual. He also realized that although people can control most of the overt behavior that is an emotional response to the situation, the internal bodily

changes necessary for that behavior and some facial and bodily movements occur so quickly that it is impossible to suppress them.

Darwin is interested only in showing how evolution can account for the expression of emotions; he is not interested in understanding the nature of the emotions. Darwin simply assumes that voluntary actions are the origin of the habitual behavior that after many generations becomes inherited and finally involves the instinctual expression of the emotion. Although I agree that Darwin's Lamarckian account is compatible with his theory of evolution by natural selection, it is not compatible with a correct account of emotions and feelings. Indeed, it even seems to be incompatible with some of the comments that Darwin makes about particular emotions and feelings when he is not putting forward a theoretical account of their origin. For example, he comments on the remark of "a child, a little under four years old, [who] when asked about what was meant by being in good spirits, answer[s], 'It is laughing, talking, and kissing.' [He says] It would be difficult to give a truer and more practical definition." (p. 210) Here Darwin seems to define "good spirits" in terms of behavior without mentioning states of mind or voluntary actions or inherited habits. Darwin is not concerned with the meaning of words. Like many scientists, he probably accepts the popular account of the meaning of referring words, that once you know what a word refers to you know what it means. This account does not create any serious problems when talking about words that refer to physical objects and properties, but when applied to terms that refer to psychological phenomena it does lead to the kinds of problems discussed in the chapters concerning private language and criterion.

Darwin seems to regard emotions, feelings, and sensations as introspected experiences, which often lead to action, but holds that even when these actions are suppressed they give rise to expressions. He says, "when movements, associated through habits with certain states of mind are partially repressed through the will, the strictly involuntary muscles, as well as those which

are least under the separate control of the will, are liable still to act, and their action is often highly expressive.” (p. 54) Darwin’s view of how emotions, feelings, and sensations explain both behavior and the expression of emotions is a very common view. It is compatible with the Lamarckian account of evolution but is incompatible with the genetic account that is now regarded as the correct view. Darwin’s lack of concern with language, including any concern about how children learn to use words that refer to psychological phenomena, results in his using the term “emotion” to refer to any state of mind that he holds responsible, directly or indirectly, for causing behavior, even if it has no natural expression. He says, “No emotion is stronger than maternal love, but a mother may feel deepest love for her helpless infant, and yet not show it any outward sign; or only by slight caressing movements with a gentle smile and tender eyes. But let anyone intentionally injure her infant, and see what a change! How she starts up with a threatening aspect, how her eyes sparkle, and her face reddens, how her bosom heaves, nostrils dilate, and heart beats; for anger, and not maternal love, has habitually led to action.” (p.83)

This quote also shows that Darwin holds that it is only the emotions that habitually lead to action that have expressions, and these expressions are those movements that accompany or precede the overt habitual behavior. Darwin is unconcerned with how anyone can learn the meaning of the phrase “maternal love, if maternal love has no natural expression.” He is right not to be concerned, for we learn the meaning of “maternal love” in the same way that we learn the meaning of all terms or phrases that refer to psychological phenomena, by using the behavior and circumstances that are their defining criterion. The defining criterion may be quite complex, with different combinations of behavior and circumstances, e.g., joy behavior when the child is pleased, sad behavior when the child is sad, anger behavior when someone hurts the child. Darwin may be confused because he makes no distinction between emotion terms and phrases when they refer to presently occurring psychological phenomena and when they refer to

dispositions to have presently occurring psychological phenomena. It is terms and phrases that refer to presently occurring psychological phenomena that are learned by means of the behavior and circumstance that are their defining criterion.

Darwin calls those terms that refer to psychological phenomena that have both behavior and circumstances as their defining criterion, “emotions,” but he also uses the term “emotion” to refer to psychological phenomena that have only behavior as their defining criterion. He refers to the former psychological phenomena as “complex emotions,” referring to the latter as “simple emotions.” Because for Darwin, simple emotions can be identified solely by their expressions; with no need to know anything about the circumstances in which they occur, I prefer to use the term “feelings” rather than “simple emotions. These psychological phenomena have behavior as their sole defining criterion; that is why they can be identified solely by their expressions. Using the term “feelings” to refer to what Darwin calls “simple emotions” emphasizes the distinction between terms that have both behavior and circumstances as their defining criterion and those that have only behavior. Although feelings have gradations, e.g., annoyance is a mild form of anger whereas rage is an extreme form; there are only a small number of feelings: anger, disgust, fear, joy, and sadness. The degree of each of these feelings is determined by the vehemence of the behavior that is their defining criterion; the behavior that is the criterion for saying that someone is enraged is far more vehement than the criterion for saying that someone is annoyed. Because all emotions incorporate feelings, my terminology allows the indefinite number of emotions to be classified in terms of the small number of feelings involved in those emotions.

Darwin and I are in substantial agreement about the distinction between emotions (complex emotions) and feelings (simple emotions), however, as we shall see, Darwin’s terminology does not make clear the conceptual difference between emotions and feelings. Darwin seems to distinguish complex emotions from simple ones in the same way that I

distinguish emotions from feeling and is properly skeptical that any psychological phenomena that he classifies as complex emotions can be reliably identified without knowing the circumstances to which the behavior is a response. However, he does not realize that there is a conceptual distinction between complex and simple emotions. He asks, “Can guilty, or sly, or jealous expressions be recognized? though I know not how these can be defined.” (p. 23) After listing the complex state of minds such as jealousy, envy, guilt, pride and humility, (Darwin also includes in this list, revenge, vanity and ambition, that I would not classify as emotions at all) he says “It is doubtful whether the above complex states of mind are revealed by any fixed expression, sufficiently distinct to be described or delineated. (p. 260) Because he does not realize the conceptual distinction between complex and simple emotions he does not have any principled objection to identifying a complex emotion solely by its facial expression. He even cites an instance in which he claims to have identified a complex emotion by its expression. “I may add, that I have observed a guilty expression, without a shade of fear, in some of my own children at a very early age. In one instance the expression was unmistakably clear in a child two years and seven months old, and led to the detection of his little crime.” (p. 261)

Although Darwin usually holds that only simple emotions can be identified by their expressions and that knowing what complex emotions are involved requires knowledge of the circumstances, he does not appreciate that there is a conceptual distinction between complex and simple emotions. That is why he is able to claim that a guilty expression sometimes can be identified solely by how it looks. He did not realize that his identification of a guilty expression depended on his knowing something about the person and the probable cause of that facial expression. He did not realize that he needed to confirm that he had correctly identified the guilty expression by finding out about the circumstances that gave rise to the expression of that emotion. A more sophisticated account of the meaning of the words that refer to emotions and

feelings would have enabled Darwin to realize that only simple emotions (feelings) can be identified solely by their expression. This is because only the terms referring to simple emotions (feelings) have behavior as their sole defining criterion. Had Darwin realized that the terms referring to complex emotions (emotions) must include both behavior and circumstances in their defining criterion, he would have realized that complex emotions (emotions) cannot be identified solely on the basis of a person's facial expression. Darwin knew that a facial expression counts as an expression of guilt only if that expression is a response to having done something guilty, but he still claimed that he had identified the expression on his child's face as a guilty expression.

When referring to presently occurring emotions, Darwin realized that internal bodily changes and identifying expressions of emotions occur before overt voluntary behavior. These internal bodily changes include physiological changes in the body as well as the tensing of various muscles in preparation for the voluntary actions that are the behavioral part of the defining criterion of those emotions. He was also aware that our internal emotional responses to circumstances usually occur so quickly that it is impossible to stop them. He realized that even when we repress the overt voluntary behavior, "the muscles which are least under the control of the will are the most liable to act, causing movements which we recognize as expressive." (p. 34) He realized that the effort to control the overt behavior often "requires other slight movements, and these are likewise expressive." (Ibid.) But prior to the slight movements that are expressive, there are other changes in the body that are not detectable. Sometimes even the person responding to the circumstances is not aware of these internal bodily changes. Darwin did not talk about what happens in the brain when we have an emotional response to circumstances because he knew very little about the brain. We now think that when presented with emotion arousing circumstances the part of the brain that controls our voluntary behavior responds by initiating the internal bodily changes that precede the overt voluntary behavior. This happens

independently of whether the part of the brain that controls our voluntary behavior is being monitored by the part of the brain that is responsible for consciousness. So we can have an emotional response, including both internal bodily changes and even overt behavior without being aware of any of this.

As explained earlier, I classify anger, disgust, fear, joy and sadness as feelings rather than as emotions, because the defining criterion for the terms referring to these psychological phenomena is simply behavior, whether or not we are aware of how we feel like behaving. The circumstances, if there are any, are not part of the defining criterion for the terms that refers to feelings. I use the term “emotion” to refer to psychological phenomena such as jealousy, pride, envy and love, because the terms referring to these psychological phenomena include both the behavioral response and the circumstances to which the behavior is a response as parts of their defining criterion. When we use an emotion term to refer to some psychological phenomena, we are claiming that it is a response to particular circumstances whether or not the person with the emotion is aware of her response or of the circumstances to which it is a response. Even if we are not aware of the emotional responses by the part of the brain that controls our voluntary behavior they are psychological phenomena because they are part of an explanation of why we act as we do. If the emotional response of the part of the brain that controls our voluntary behavior is not being monitored by the part of the brain responsible for consciousness we will be completely unaware of how we feel like acting and sometimes even of our overt voluntary behavior.

I have tried to explain why I use the term “feelings” to refer to anger, disgust, fear, joy and sadness, even when we are not aware that we have these feelings. Although we are often aware of these feelings it is not uncommon for us to have these feelings without being aware of them, that is, we can be unaware that we are about to behave in the way that serves as the defining criterion for a term such as anger. I realize that, since Freud, it is possible to talk about

unconscious feelings, but I do not want to limit “feelings” to those psychological phenomena that either we are aware of or that are unconscious in the Freudian sense. I mean by “feelings” those psychological phenomena for which voluntary behavior is the defining criterion, regardless of whether the person having those feelings is aware or conscious of them. To say “anger,” “disgust,” “fear,” “joy” and “sadness” refer to feelings only when one is aware of how one is about to behave would require using different terms to refer to anger, disgust, fear, joy and sadness when we are not aware that we are about to behave in ways that are the defining criterion for the terms that refer to these feelings. This would make it impossible for someone to say that he had just become aware that he was angry. To say that a person has become aware that he is angry is to say that he has become aware of how he is about to behave, or that he has become aware of how he feels like behaving.

I use the phrase “bodily sensations” to refer to the internal bodily changes, e.g., increased heart rate or tensed muscles, that precede overt voluntary behavior. Bodily sensations are those internal bodily changes that can be part of an explanation of why I do some voluntary action, regardless of whether I am aware of them. All feelings are psychological phenomena because they can be part of an explanation of why I do some voluntary action regardless of whether I am aware of them. Not all bodily changes are psychological phenomena. Only those bodily changes that can be part of an explanation of why I do some voluntary action are bodily sensations, but these bodily changes are sensations regardless of whether I am aware of them or whether they lead to a voluntary action in this particular situation. Internal bodily changes of which one is aware are always bodily sensations, but internal bodily changes of which we are unaware are bodily sensations if they can be part of an explanation of why I do some voluntary actions.

We become aware of our feelings when the part of the brain responsible for consciousness is monitoring the part of the brain that controls our voluntary behavior but this

awareness is not the cause of the overt behavior that normally results from these feelings. Insofar as the awareness of our feelings has any effect on our overt behavior it is that being aware of how we are about to behave enables us to suppress the overt behavior that would otherwise occur. We suppress the overt behavior when we believe that it would be inappropriate or otherwise harmful to act as we feel like acting. In fact, when a person acts as he feels like acting when it is inappropriate to act in that way, he is sometimes said to be “acting out” and it is regarded as the result of a lack of insight or a lack of control. Being aware of our feelings is not being aware of various internal bodily changes, e.g., tensing muscles or increased heart beat; being aware of these internal bodily changes is being aware of our bodily sensations. Being aware of our feelings is also not being aware of what is going on in our brain in any normal sense. Except for those experimenting on themselves, people are not aware of what is going on in their brain. However, when the part of the brain responsible for consciousness is monitoring the part of the brain that controls our voluntary behavior we become aware of how we are about to behave or how we feel like behaving. Feeling angry is feeling like hitting; feeling disgust is feeling like spitting something out of one’s mouth; feeling fear is feeling like running away; feeling sad is feeling like crying; and feeling joy is feeling like smiling even if one is not aware that one is about to behave in that way. (I use “joy” instead of “happiness” or “pleasure” primarily because this is the term that Darwin and Hobbes use to refer to the presently occurring positive feeling or simple emotion.)

Darwin realizes that the circumstances to which some overt behavior is a natural response is often very good evidence that one has the feeling that leads to that overt behavior. However, one need know nothing about the circumstances to which one is having an emotional response in order to identify the feeling. Indeed, one of the main points of Darwin’s book is to show that the five feelings: anger, disgust, fear, joy and sadness can usually be identified solely on the basis of

their facial expressions. However, because Darwin follows Spenser in using “feelings” as a general term that refers to both sensations and emotions he does not have a distinct category of feelings. (p. 33) Nonetheless, Darwin does distinguish anger, disgust, fear, joy and sadness, from complex emotions such as pride and jealousy, and calls them simple emotions. Classifying anger, disgust, fear, joy and sadness as feelings rather than classifying them as “simple emotions” provides a clearer way to group the complex emotions into categories. What Darwin calls complex emotions such as pride and jealousy can be grouped according to the feelings that they involve; pride involves joy and jealousy involves anger. Terms that refer to complex emotions, or what I simply call emotions, such as “pride” and “jealousy,” include both behavior and circumstances as their defining criterion, whereas terms that refer to feelings only have behavior as their defining criterion. Having the category of feelings allows jealousy and indignation to be grouped together as both involve the feeling of anger; they differ primarily in the circumstances to which the behavior is a response, although the anger is also exhibited in different ways.

Nonetheless, anger, disgust, fear, joy and sadness are generally behavioral responses to general kinds of circumstances, e.g., anger to frustration or thwarting of desires or plans, disgust to something completely unacceptable, fear to danger, joy to doing what one wants to continue doing, and sadness to loss. People so commonly react to thwarting of desires with hitting, to something disgusting with spitting something out, to danger with fleeing, to continuing to do what one is doing with smiling and to loss with crying, that we regard these behaviors as natural ways to respond to these circumstances. The relationships between these general kinds of circumstances and the behavior that is a natural response to them can be regarded as principles of human nature. These relationships are so strong that we sometimes even use circumstances to describe these feelings. For example, saying “I feel as if my best friend just died,” would be taken as describing sadness; saying “I feel as if my plans are being sabotaged” might be taken as a

description of anger; saying “I feel like I some muggers are following me down a dark alley” would be taken as describing fear; saying that I feel like something slimy is in my mouth would be taken as describing disgust; and saying “I feel as if the most wonderful thing just happened” would be taken as describing joy. Darwin did not make a mistake by classifying anger, disgust, fear, joy and sadness as simple emotions, for they are usually responses to circumstances, but calling them feelings emphasizes Darwin’s point that only simple emotions can be identified by facial expressions; no knowledge of the circumstances is needed. Facial expressions serve to identify particular feelings because of the very close relationship between these expressions and the overt voluntary behavior that is the defining criterion of the term referring to the feeling.

The feelings of anger, disgust, fear, joy, or sadness are not always part of an emotion, i.e., they are not a response to any circumstances. When we are aware of how we feel like behaving but know that feeling this way is not part of an emotional response, we may use different words to refer to how we feel like behaving, even when we feel like behaving in the same way we do when we feel anger, disgust, fear, joy, or sadness. So, instead of using the words, “anger,” “fear,” and “sad” we say, “I feel irritable,” “I feel anxious,” or “I feel depressed.” When we are talking about a presently occurring feeling, the defining criterion for feeling irritable is in the same series of behaviors as the defining criterion for feeling anger or annoyance. Similarly, when we are talking about a presently occurring feeling, the defining criterion for feeling anxious is in the same series of behaviors as the defining criterion for feeling various degrees of fear. Likewise, when we are talking about a presently occurring feeling, the defining criterion for feeling depressed is in the same series of behaviors as the defining criterion for feeling various degrees of sadness. There is a wide range of behaviors that can serve as a defining criterion of a feeling and I do not hold that there need be a specific behavior that is the defining criterion of a feeling.

It is quite common to talk of someone getting more or less annoyed or angry, or becoming more or less afraid or sad. The intensity of the feeling varies with the intensity of the behavior that is the defining criterion of the feeling. The behavioral defining criterion of “rage” will be far more violent than the behavioral defining criterion of “annoyed,” but they will belong to the same series. A reason for saying that the behavior that is the defining criterion of “rage” is in the same series as the behavior that is the defining criterion of “annoyed” is that they are responses to circumstances that vary in degree. For example, if your car won’t start after a few tries, you are likely to manifest the behavior that is the defining criterion of “annoyed,” but if you have an important appointment, and the car doesn’t start after several more tries, you are likely to manifest the behavior that is the defining criterion of “anger,” and several more unsuccessful attempts may lead one to manifest the behavior that is the defining criterion of “rage.” Similarly, as the magnitude of the danger one faces becomes clearer the behavior that one displays will change, so that the behavior that was the defining criterion of “fearful” may become behavior that is the defining criterion of “terror.” Degrees of sadness and joy, which often have different names, vary with degrees of loss or gain, the latter situation being commonly provided by quiz shows as they slowly reveal the better and better prizes won by the contestants. The change in behavior of the contestants as this happens is naturally considered to be the same kind of behavior only more intense.

Darwin uses evolutionary arguments to explain why there are universal facial expressions involved in what he calls the simple emotions, viz., anger, disgust, fear, joy and sadness. Darwin holds that facial expressions are explained by the same bodily changes that lead to overt voluntary behavior. He claims that these universal expressions of the simple emotions or feelings are a prelude to the natural intentional responses to some general kinds of circumstances. There is no longer any doubt that Darwin was correct in identifying the facial expression of fear by

noting its association with the fearful behavioral response to a dangerous situation. Without completely realizing what he was doing, Darwin identified fear by using its behavioral defining criterion and the dangerous circumstances to which this behavior is a response. That was what enabled him to conclude that the expression of fear is universal. The same is true of the facial expressions of anger, disgust, joy, and sadness. This strongly suggests that the current emotional behavioral responses to danger, frustration, disgusting food, gain, and loss were generally advantageous to the animals that had those responses. Thus, without invoking intentional actions that become habits of responding, and are eventually inherited, we can account for the relationship between certain kinds of circumstances and corresponding kinds of voluntary behavior. This genetic account explains the universality of the expressions of the feelings as well or better than Darwin's Lamarckian account.

What Darwin calls complex emotions, and I simply call emotions, such as pride and jealousy, cannot be identified solely by behavior and facial expressions; rather the defining criterion must involve a combination of behavior and the specific circumstances that the behavior is a response to. One can know that a person is angry without knowing what, if anything, made him angry, although it is often assumed that the circumstances involve something happening that frustrated the person's plans. But anger can be identified as jealousy only when the circumstances are such that one knows that the person believes that the exclusive loving relationship that he valued may have been lost because of the loved one's relationship to someone else. The circumstances that are part of the defining criterion of jealousy are combined, especially in men, with the behavioral criterion of anger. It is not uncommon for a child basking in the love of his mother to become angry when his mother stops playing with him to respond to the call of a sibling. The child may then be told, not to be so jealous and so the child learns that being angry at the disruption of an exclusive love relationship is the defining criterion of "jealousy." However,

some women claim that sadness is also a natural response to the circumstantial defining criterion of jealousy, and many people seem to hold that fear is also a natural response to the circumstantial defining criterion of jealousy. However, English has the expression “jealous rage” but not the expressions “jealous panic” or “jealous depression” and this suggests that angry behavior, which is what children exhibit, is the natural behavioral defining criterion of jealousy.

Using what we now regard as the correct Darwinian account of evolution by natural selection we arrive at a very different picture of relationship between (1) states of mind, especially conscious states of mind, and (2) appropriate or serviceable behavior, and (3) the expression of emotion. Darwin is right about the relationship between (2) appropriate or serviceable behavior, and (3) the expression of emotion. As this is his primary concern, much of what he says in his book on the explanation of the expressions of the emotions is correct. We recognize the expressions of the feelings, what Darwin calls the simple emotions, of anger, disgust, fear, joy and sadness, because these expressions precede the behaviors that turned out to be beneficial responses to the kinds of situation that provoked them. We may know which behaviors that we take as behavioral defining criterion for anger, disgust, fear, joy and sadness, but may not know why these responses turned out to be more beneficial than alternative responses. However, if there is an explanation for why certain kinds of circumstances, e.g., dangerous ones, almost universally lead to certain kinds of fear responses, e.g., running away, the doctrine of natural selection provides that explanation. Once we accept that these behavioral responses increased the probability of the animals’ survival then we have an explanation of why the expressions that precede these kinds of behavior are the natural expressions of emotional behavioral responses to those circumstances.

The fear behavioral response to danger seems so obviously appropriate that it may have led Darwin into accepting the Lamarckian view that in the beginning there was an intentional

voluntary action to avoid the danger and that later this became a habit and finally became inherited as an instinct or reflex. However, the now standard view is that animals that had mutations that resulted in their responding to dangerous situations with avoidance or other fear behavior such as freezing or hiding, survived and reproduced at a higher rate than those that did not have the mutations that led to these kinds of responses. With regard to fear responses, it seems obvious that they are beneficial responses to danger. However, it is not obvious that sadness behavior is an advantageous response to what we now regard as situations of loss. Accepting that natural selection must be the explanation prompts us to look for reasons why mutations that resulted in sadness behavior were beneficial. The following is a just-so story that can provide an explanation for why sadness behavior is beneficial in circumstances of loss. Taking the loss of a parent as the most significant loss that an infant of almost all mammalian species would experience, then it is plausible that responding to these circumstances by staying where one is, not moving, etc., may be more beneficial for the survival of the child than any other response.

Expressions of joy, which may seem quite violent, even though they are not followed by what even seems to be purposeful behavior, may be an example of a modified version Darwin's second principle. Since joy may be followed by behavior that could be taken as anger behavior, the facial expression of joy may prevent misunderstanding and hence not provoke a hostile response to that behavior. Smiles may reassure those watching the violent looking behavior that follows that this behavior is not anger behavior, and so might be beneficial not only in increasing the chances of survival, but also increasing the chances of engaging in reproduction. Smiles seem to be easily recognizable from great distances, and even though smiles that are part of joyful responses were not originally done intentionally to have beneficial consequences, a genetic explanation both of smiling and of responding appropriately is quite plausible.

Taking natural selection of certain behaviors as an explanation for why we now take these kinds of behavior as defining criterion for certain feelings, e.g., fear behavior for fear and sadness behavior for sadness, we arrive at a view very different from the view that Darwin accepted about the nature of emotions and feelings. On that view, states of mind or experiences, lead us to act in an appropriate way to the circumstances, e.g., the state of mind of fear causes us avoid danger, and this eventually becomes a habit and then is inherited as an instinct or reflex. On the view I am presenting, there are genetic mutations that lead to differing behavioral responses to danger; those mutations that lead to the behavioral responses that we now call fear behavior, result in the animals with those mutations surviving at a greater rate than those with different mutations. States of mind or intentional actions are completely irrelevant to this process; all that is relevant is that some kinds of behavioral response to danger, e.g., running away, have greater survival value. This kind of behavioral response, fleeing from danger, we call fear behavior and we use that behavior as a defining criterion of fear. Note that this explanation can account for the behavior and expressions for all animals that exhibit the kind of behavior that we use as defining criterion for fear, without even mentioning states of mind as causes of those actions or expressions. Contrary to Darwin's Lamarckian account, the genetic account of evolution, does not require any conscious state of mind or intentional action to develop into a habit that then becomes an inherited instinct or reflex.

In the Introduction to *The Expression of the Emotions in Man and Animals*, Darwin noted, "The study of expression is difficult owing to the movements being often extremely slight, of a fleeting nature." (Page 12.) Anticipating Freud, Darwin decides to observe infants, "for they exhibit many emotions, as Sir C. Bell remarks, 'with extraordinary force:' whereas in after life, some of our expressions 'cease to have the pure and simple source from which they sprang in

infancy.” (P 13) and “that the insane ought to be studied, as they are liable to the strongest passions, and give uncontrolled vent to them.” (Ibid.) He also showed photographs of an old man whose facial muscles had been electrically stimulated, “without a word of explanation, to above twenty educated persons of various ages and both sexes, asking them, in each case, by what emotion or feeling the old man was supposed to be agitated.” (Page 14) “Several of the expressions were instantly recognized by almost everyone . . . On the other hand, the most widely different judgments were pronounced in regard to some of them. (Ibid.) Darwin regards the former, about which there was almost complete agreement, i.e., anger, disgust, fear, joy and sadness as being the natural expression of a simple emotion.

Finally, in order to see “whether the same expressions and gestures prevail, . . . with all the races of mankind, especially with those who have associated but little with Europeans.” (Page 15) Darwin sent a list of 16 questions to correspondents in all parts of the world. He was quite clear about the point of this study. “Whenever the same movements of the features or body express the same emotions in several distinct races, we may infer with much probability, that such expressions are true ones. – that is, are innate or instinctive. Conventional expressions or gestures, acquired by the individual during early life, would probably have differed in the different races, in the same manner as do their languages.” (Ibid.) It is instructive to list some of the questions that Darwin sent out.

1. Is astonishment expressed by the eyes and mouth being opened wide, and by the eyebrows being raised?

10. Is disgust shown by the lower lip being turned down, the upper lip slightly raised, with a sudden expiration, something like incipient vomiting, or like something being spit out of the mouth?

11. Is extreme fear expressed in the same manner as with Europeans?

15. Can guilty, or sly, or jealous expressions be recognized? though I know not how these can be defined.

16. Is the head nodded vertically in affirmations, and shaken laterally in negation?

In asking some of these questions, Darwin uses circumstances and behavior to help identify the feeling or emotion and then tries to discover whether there is an empirical connection between the emotion or feeling so identified and a given expression. Darwin does not make Hume's mistake of thinking that he can identify the feeling or emotion by introspection. Rather, he holds the much more plausible view that given certain circumstances and behavior a given feeling or emotion will be felt, and he is interested in knowing whether this feeling or emotion is universally expressed by the same expression. The following remark shows how different his view is from that of Hume. "I heard a child, a little under four years old, when asked what it meant to be in good spirits, answer, 'It is laughing, talking, and kissing.' It would be difficult to give a truer and more practical definition" (page 210)

What Darwin's regarded as important about his account of the relationship between expressions and emotions does not involve any claims about the source of the changes that were subject to natural selection, i.e., whether they were came about genetically or by inheritance. But he was certain that evolution by natural selection was involved. Darwin says, "No doubt as long as man and all other animals are viewed as independent creations, an effectual stop is put to our natural desires to investigate as far as possible the causes of Expression . . . He who admits on general grounds that the structure and habits of all animals have been gradually evolved, will look at the whole subject of Expression in a new and interesting light." (Op cit. page 12.) Darwin shows that when specific circumstances naturally elicit specific kinds of behavior, the body in preparing to behave in that way has certain kinds of physiological changes. He shows that these preparations for intentional behavior would naturally show themselves in what we now regard as

an expression of the feeling or emotion. His major concern was to deny that expressions of emotions were created in order to serve as signals or signs to others. He did not deny that the naturally occurring expressions could come to have beneficial consequences by serving as signs to other animals so as to avoid a misinterpretation of the behavior that followed.

In the Introduction to *The Expression of the Emotions*, (page 10), Darwin says, “But the simple fact that the anthropoid apes possess the same facial muscles as we do, renders it very improbable that these muscles in our case serve exclusively for expression; . . . Distinct uses, independently of expression, can indeed be assigned with much probability for almost all the facial muscles.” He argued against the view that the expressions of emotions were developed for purposes of communication because that would suggest that the expressions of human beings did not evolve from the expression of animals. Darwin did not deny that expressions came to be recognized by others as expressions of specific feelings or emotions, but he did not think that this explained their origin. If the origin of an expression were as a signal or sign, it would be completely implausible that in every culture and society, the expression of a specific feeling or emotion was always the same. Whereas, if the origin of an expression could be explained as deriving from internal changes that resulted from the emotional response to the circumstances and that lead to the subsequent behavior, then it would be natural that the expression would be the same. So, Darwin held, at least implicitly, that certain kinds of circumstances naturally give rise to certain kinds of behavior, and hence to certain kinds of expressions. It would not be surprising that these expressions would soon be taken as signs or signals of the circumstances and behavior, at least in a general way. It is not incompatible with Darwin’s account that certain kinds of expressions are now regarded as essential features of a feeling or emotion.

Darwin was not primarily, if at all, interested in explaining the meaning of the words that refer to feelings and emotions, but he was an extraordinarily perceptive observer. Thus it is not

surprising that he noticed the way words were used and the conditions under which we attributed emotions and feelings to other persons. He was so taken by the close connection between the criterion for the terms that referred to feelings and emotions that he sometimes even defined those terms by using the criterion, as in the quote of the little girl cited above. For all of those feelings that can be identified by facial expressions, e.g., anger, disgust, fear, sadness, and joy, Darwin simply assumed that there are some circumstances that naturally gave rise to those feelings. Anger would be a natural response to the thwarting or frustrating of a desire, disgust to spoiled food, fear to danger, sadness to loss, and joy to continuing satisfaction of desires.

These responses to the kinds of circumstances are so obvious, that to claim that they illustrate some principles of human nature makes it seem as if these principles must be analytic. How could it merely be an empirical fact that fear is a natural response to danger, or sadness to loss? It is not merely an empirical fact, but it does presuppose an empirical fact. Danger is the circumstantial part of the defining criterion of “fear” when it is taken as referring to an emotion, and loss is the circumstantial part of the defining criterion of “sadness” when it is taken as referring to an emotion. Behavior and expressions are the behavioral part of the defining criterion of “fear” and “sadness” when they refer to emotions, and they are the sole defining criterion when they refer to feelings. Every language has words that refer to fear and sadness, whether we take these terms to refer to emotions or feelings. The facial expressions of fear and sadness are universal because “fear” and “sadness” have universally the same behavioral defining criterion. And it is also universal that their defining criterion arise from the same kind of circumstances; danger for fear and loss for sadness.

If the term for a complex emotion, e.g., shame, is universal that presupposes that there is a common and natural relationship between certain kinds of circumstances, e.g., having a characteristic that one regards as making one inferior, and certain kinds of behavioral responses,

e.g. sadness behavior. If this relationship is built into all languages, that is, if there is a word in every language that has the same circumstantial and behavioral criterion, then it might seem as if the connection between the circumstances and the behavior is simply linguistic or analytic. But that all languages have a term with the circumstantial and behavioral defining criterion of shame shows a feature of human nature, namely that acting in ways that are the behavioral criterion of sadness is a natural response to believing that one is inferior. Perhaps this is a development of submissive behavior in some hierarchical mammal species, but regardless of its evolutionary origin, the connection between regarding oneself as inferior in some way and showing sadness behavior seems to be universal. This universal relationship demonstrates what we all know, but do not take seriously enough, that our languages depend on the world being the way it is. This is why we can sometimes make discoveries about the world from examining language.

Accounting for introspectable psychological phenomena

From what I have said so far in this chapter, it seems as if emotions and feelings do not involve introspectable psychological phenomena at all. However to claim that introspectable psychological phenomena are never involved with emotions and feelings would be to hold the behaviorist account of psychological phenomena that I criticized in the very first chapter. What I have shown in this chapter is that states of mind, or introspectable psychological phenomena, are not needed to explain behavioral responses to circumstances or states of the body. On the view that I am putting forward, introspectable psychological phenomena evolved when circumstances changed more quickly than the behavioral responses to circumstances could evolve. In changing circumstances, whether caused by moving to a different environment, or significant changes in one's present environment, behavioral responses to circumstances that were beneficial may cease to be beneficial and may actually become harmful. Beneficial mutations do not occur frequently enough to enable the behavioral responses to adapt to rapidly changing circumstances.

At some point in evolutionary history, whether with the development of mammals, or of primates, or of Homo sapiens, genetic mutations that enabled animals to modify their natural behavioral responses to circumstances became sufficiently beneficial that those who inherited those mutations survived at a greater rate than those without those mutations. I am going to tell a just-so story about what these mutations involved. Of course, these mutations did not occur all at once, but I am not concerned with the details of their development, all that is important is that a series of mutations enabled the animals that had them to modify their behavioral responses to circumstances. These mutations had to involve at least two new abilities of which the first was to become aware of how they were about to behave before they actually behaved in that way. This mutation involved the part of the brain responsible for consciousness monitoring the part of the brain that controlled one's voluntary behavior. When that happens the person is aware of how feels like behaving or how he is about to behave. Knowing that one's emotional response will lead one to behave in a particular way is like knowledge of one's intentions; it is not sensory knowledge, i.e., the kind of knowledge that comes from the senses, e.g., eyes, ears, skin, etc. Knowledge of the states of one's body is sensory knowledge and is like knowledge of the external world that is also the kind of knowledge that comes from the senses. Knowing how one is about to behave is different: it provides the flexibility of behavior that is beneficial when one's natural behavior no longer reliably results in good consequences in a changed environment.

In addition to this ability to know how they are about to behave if they do not suppress their natural emotional behavioral response, they must also have had mutations that enabled them to distinguish between the consequences of their natural response and the consequences that they want to occur. Distinguishing between the natural consequences of their behavior and the consequences they want is the beginning of distinguishing between means and ends. The ability to know how one is about to behave will not have a beneficial effect on one's behavior unless

one is also aware of the consequences of behaving in that way, and realizes that some other consequences are more beneficial than the consequences of behaving as one feels like acting. This ability, like the previous ability, must have involved some changes in the brain; perhaps the ability to compare the consequences of different courses of action. Animals, including human beings, sometimes act on instinct or impulse without thinking about the consequences of their actions. Indeed, as I have been maintaining throughout this chapter, most actions that we regard as acting on our emotions or feelings are primarily the actions that are the natural outcome of the brain responses to circumstances that involve no conscious thought at all, let alone conscious thought about the consequences of that behavior. It is the beginning of maturity in children that when they are aware of how they are about to behave, they change or modify their natural responses if they realize that the consequences of their natural responses would be harmful to them. Acting this way is acting on reasons, e.g., on beliefs about harms to be avoided in the future, rather than acting on their emotions, i.e., simply responding emotionally to circumstances.

It is important to note here that the view of emotions and feelings as behavioral responses to circumstances or the state of one's body does not deny that people are often aware of their feelings and emotions in a way that no one else can be aware of them. And although this awareness is not the cause of the natural behavioral response, it is not merely an epiphenomenon; for this awareness enables the person to suppress the natural behavioral response. To state the matter in a paradoxical sounding way: the awareness of how one is about to behave comes after one has already begun the internal bodily changes that precede the natural emotional behavioral response; it is not the cause of that natural emotional behavioral response. However, this awareness may result in suppression of the natural behavioral response. Paradoxically, the awareness of how one is going to behave has a causal role only when it results in one not acting in that way. Emotional behavior is not goal driven, that is, it is not driven by goals to be

achieved, but rather it is impulse or instinct driven. It is caused not by beliefs about the future, but rather by reactions to what is present or past. It is only when awareness of how one is about to behave results in restraint of that behavior that it is appropriate to talk about rational or goal-oriented behavior.

This biologically based account of emotions and feeling confirms the linguistically based account presented in the previous three chapters. In those chapters, the meaning of the terms that refer to emotions and feelings is determined by behavioral and/or circumstantial criterion. The defining criterion for feelings is as follows: for “anger” is anger behavior; disgust, disgust behavior; for fear, fear behavior, for joy, joy behavior; and for sad, sadness behavior. I have already given some descriptions of what counts as anger behavior, disgust behavior, fear behavior, joy behavior, and sadness behavior, but anthropologists, biologists, and psychologists are much better at doing this than I am. The defining criterion for emotions includes both behavior and circumstances: for jealousy, it is anger behavior and the disruption of an exclusive love relationship because of the loved one’s relationship to someone else; for pride it is joy behavior and the accomplishment of a significant task; for grief it is sadness behavior and the death of a close family member. The introspected feeling must be described in term of the behavior that is the criterion of the term that refers to the feeling: e.g., feeling fear is feeling like fear behaving, feeling joy is feeling like joy behaving. The introspected emotion is the result of the part of the brain responsible for consciousness monitoring the part of the brain that controls our voluntary behavior. The description of this conscious emotion must use the behavior and circumstances that are the defining criterion of the term that refers to the emotion. For example, feeling pride is feeling like joy behaving because of accomplishing a significant task, feeling jealousy is feeling like behaving with hostility toward the person with whom one had an exclusive love relationship because of her love relationship with someone else.

Given the current understanding of evolution, those behavioral responses to circumstances that resulted in a higher probability of survival and/or reproduction became the normal behavioral response of members of that species. We can now identify anger responses, disgust responses, fear responses, joy responses, and sadness responses without even mentioning states of mind, intentions, goals, or even consciousness. These behavioral responses are accompanied by or preceded by facial expressions that are so distinctive that seeing the facial expression enables one to know the kind of behavioral response to which it is related even when one does not know any details about the circumstances to which the behavior is a natural response. I have tried to make clear that I use the term “feelings” to refer to those psychological phenomena for which behavior is the sole defining criterion. To feel like behaving in some way is to be about to behave in that way and it is not necessary that we be aware or conscious of feeling like behaving in that way, i.e., of being about to behave in that way. Even when we are aware of or conscious of feeling like behaving in a specific way, e.g., running away, the consciousness of the feeling is not a cause of the behavioral response. Insofar as the awareness or consciousness plays a causal role, it is when we want to suppress our natural emotional behavioral response. Being conscious of how one feels like behaving is beneficial in circumstances when the environment has changed so much that being able to suppress one’s natural behavioral response increases the probability of survival and/or reproduction.