Looking into meta-emotions

Christoph Jäger · Eva Bänninger-Huber

Abstract There are many psychic mechanisms by which people engage with their selves. We argue that an important yet hitherto neglected one is self-appraisal via meta-emotions. We discuss the intentional structure of meta-emotions and explore the phenomenology of a variety of examples. We then present a pilot study providing preliminary evidence that some facial displays may indicate the presence of meta-emotions. We conclude by arguing that meta-emotions have an important role to play in higher-order theories of psychic harmony and that Frankfurt-style accounts, which explain a person's "reflective self-endorsement" exclusively in terms of *volitional* hierarchies, are inchoate and need to be augmented by a theory of meta-emotions.

Keywords Meta-emotions · Higher-order emotions · Higher-order desires · Psychic unity · Self-harmony · Harry Frankfurt · Paul Ekman

1 Introduction

"The only thing we have to fear is fear itself", Franklin Roosevelt declared in his inaugural address (1933). However we may interpret and evaluate the political implications of this statement, it relies on an interesting philosophical and psychological assumption. Not only do we have emotions; we also have emotions about our emotions. Let's call such higher-order affective states or episodes "meta-emotions".

C. Jäger (✉)
Department of Philosophy, University of Innsbruck, Karl-Rahner Platz 1, 6020 Innsbruck, Austria
e-mail: christoph.jaeger@uibk.ac.at

E. Bänninger-Huber
Department of Psychology, University of Innsbruck, Bruno-Sander-Haus, Innrain 52f, 6020 Innsbruck, Austria

Despite their prevalence and, we maintain, their great significance for understanding the mechanisms by which people engage with their selves, to date meta-emotions have not received much systematic attention in the literature. This article begins to fill this lacuna. We analyze the core features of meta-emotions, explore their phenomenology, and argue that they play a pivotal role in determining psychic harmony.[1]

Section 2 distinguishes various kinds of meta-emotion and discusses their intentional structure. Yet some may remain skeptical. In particular, empirically minded philosophers might wish to see whether our claims can be implemented empirically.[2] We have risen to this challenge. In Sect. 3 we present and discuss a pilot study which provides preliminary evidence suggesting that some facial expressions may indicate the presence of meta-emotions, and that people, when striving for self-harmony, engage in meta-emotional appraisals of their affective lives. Section 4 outlines the significance of our overall argument for the philosophy of the self. Among the most influential theories of psychic harmony are Frankfurt-style accounts which analyze the relevant form of self-appraisal in terms of *volitional* hierarchies. However, to begin with, the relevant first-order objects of self-endorsement or of distancing oneself from one's self include not only desires but also emotions. Moreover, not only do we relate to our first-order selves by having higher-order desires or volitions; we also engage in emotional higher-order self-appraisals. If, as most theories of the emotions maintain, emotions cannot be reduced to desires, it follows that volitional accounts of psychic harmony need to be supplemented by a theory of meta-emotions.

2 Meta-emotions: varieties and core characteristics

In Roosevelt's dictum, a hedonically negative higher-order emotion (fear) is directed toward a negative first-order emotion of the same type (fear). But negative emotions also combine with negative emotions of different types, as when we are ashamed (angry, annoyed) about feeling afraid, or afraid of being ashamed (angry, annoyed). They also come in positive-positive combinations. Consider an example from Aquinas: Concurring with Roosevelt that we can fear fear, Aquinas points out that one can also

---

[1] Among the few recent articles that have addressed meta-emotions from a philosophical point of view are Jäger and Bartsch (2006) and Mendonça (2013). Jäger and Bartsch (2006, p. 199) conclude that "whatever format an acceptable theory of emotions adopts, it should be equipped to deal with meta-emotions". I stick to this claim. Mendonça (2013, p. 390) concurs with us about the "necessity of considering metaemotions for a complete emotion theory", yet at the same time she argues that "metaemotions cannot be handled as a special case of emotion because reflexivity modifies the nature of our emotional world" (p. 391). Even if this latter claim were true we cannot see how it would follow that meta-emotions fail to constitute emotions. And if they did, why should a theory of the emotions account for them? A brief discussion of "layered emotions" can also be found in Pugmire (2005), p. 174, who argues that "certain types of emotion themselves reverberate in the mind emotionally". Susan Feagin has argued that the "pleasures of tragedy" are "meta-responses" arising from "our awareness of, and in response to, the fact that we do have unpleasant direct responses to unpleasant events as they occur in the performing and literary arts" (1983, p. 209).

[2] Philosophers who stress this include Griffiths (1997), Prinz (2004), Robinson (2004), De Sousa (2010), cf. p. 96), and Nussbaum (2013). For example, Griffiths (1997, p. 1) writes that "questions about the nature of emotions cannot be answered in the armchair alone but must be sought in part by empirical investigation of emotional phenomena".

suffer from "sorrow about [a given episode of] sorrow" and that "man can love love".[3] People may also feel happy about falling in love, or joyful about being happy. Think of the overambitious artist who is notoriously dissatisfied with his work. Yet one day he produces a piece that pleases him and feels happy not only about his artistic achievement but also about experiencing this first-order happiness.

In yet other (perhaps psychologically more complex) cases the hedonic qualities of the higher- and the lower-level emotions differ. Typical positive-negative instances are first-order emotions with an intrinsically negative hedonic tone which their subjects nevertheless experience as positive. When in Henry James's *The American* Christopher Newman finally comes to realize that his beloved Claire de Cintré is forever lost to him, he falls into a state of deep melancholy. Yet, we are told, "the dullness of his days pleased him; his melancholy, which was settling into a secondary stage, like a healing wound, had in it a certain acrid, palatable sweetness" (James 1876/77, chapter 26, opening passage). Similar phenomena are portrayed in a great many other famous works of literature. (Nineteenth-century fiction appears to be particularly rich in this respect.)[4]

Finally, consider negative emotions that are directed at positive ones. "I hate how much I love you", is a line from a recent pop song (performed by US singer Rihanna) that springs to mind. It may not be entirely clear exactly what Rihanna means by "love" here. In any case, if you are disinclined to count love as an emotion, other examples are readily available. Consider feeling bad about your *Schadenfreude*. Despicable as it may be, experiencing pleasure or amusement at another's misfortune is a hedonically positive emotion; hence feeling ashamed (guilty, disappointed, or angry at yourself) that such an emotion gained hold of you is an example of entertaining a hedonically negative emotion about a hedonically positive one.[5] To summarize our initial observations: Reflexive engagement with one's self often appears to involve higher-order affective states or episodes, which may occur in negative-negative, positive-positive, positive-negative, or negative-positive pairings.

There are various debates in the philosophy of the self which we believe could profit from considering meta-emotions. In Sect. 4, we will relate our discussion to theories of psychic harmony. Another example is the phenomenon of so-called "mixed emotions". According to traditional analyses, having mixed emotions is a complex psychic attitude combining positive and negative emotions about the same object or situation. This is typically explained in terms of different aspects under which the

---

[3] "Homo potest amare amorem, et dolere de dolere. Ergo etiam pari ratione potest timere timorem" (Aquinas, ST, I–II, q. 42, art. 4, s.c. and *corpus*).

[4] "Melancholy is the joy of being sad", says Victor Hugo ("La mélancolie, c'est le bonheur d'être triste", *Les Travailleurs de la Mer*, part III, book 1, chapter 1, p. 252.). In Tolstoy's *War and Peace* Count Pierre Bezukhov feels devastated when he has split up with his wife and killed his suspected rival in a duel: "Everything within and around him seemed confused, senseless, and repellent. Yet in his very repugnance of his circumstances, Pierre found a kind of tantalizing satisfaction" (book 5, chapter 1). Consider also a passage in Augustine's *Confessions*, when he contemplates the seemingly paradoxical fact that sometimes "sorrow itself becomes a pleasure" and "tears and sorrows are loved" ("[D]olor ipse est voluptas eius. ... Lacrimae … amantur et dolores." (III, 2, 2, and 2, 3.)

[5] For an interesting negative-positive example see also David Pugmire (2005, p. 174): "I may be … aghast to find that I am relieved at a certain thing (that a defining personal challenge has passed me by)".

situation is experienced. For example, the classic story that you can feel both good and bad when a friend obtains something you wanted as well (a position you had both applied for, for example) is analyzed in terms of mixed feelings attributable to the fact that the situation is experienced as having both positive and negative aspects. However, often things are more complicated than this. You may not only feel happy about your friend's achievement and disappointed about your own failure, but may also be dissatisfied with your happiness or feel guilty for your disappointment. Moreover, often ambivalent emotions seem hard to explain in terms of both positive and negative aspects of one and the same intentional object. Consider righteous indignation. Often, people appear to be scandalized by someone else's behavior and at the same time to experience some kind of satisfaction or even joy related to the situation. Typically, however, these latter feelings don't pertain to any positive aspects which the subject attributes to the intentional object of her indignation. A much more plausible explanation is that she has positive feelings *about her indignation*, perhaps because she regards this latter emotion as morally appropriate or as indicating her own moral superiority. In general, we believe that, even if the existence of meta-emotions were uncontroversial, it is profitable to explore their phenomenology and intentional structure, thereby providing a basis for applying the concept to various questions from the philosophy of the self and the theory of emotions.

Reductive judgmentalist accounts, as famously suggested by the Stoics and contemporary emotion theorists such as Robert Solomon or Martha Nussbaum, may construe meta-emotions as just some kind of higher-order evaluative judgment or belief.[6] Other approaches (which we regard as more promising) break down the relevant intentional relations into several components, typically including at least an informational element, an evaluative or "valence dimension", and a motivational element.[7] Theories from the latter family might say that, when a person has a meta-emotion, the information she receives (via belief or perception) is about some of her first-order emotions. She then assesses this information with respect to her goals or preferences, thereby creating a motivation either to maintain or to change the emotion-eliciting situation. In the present case, the emotion-eliciting situation is not some external affair, but one of the subject's first-order emotions.

Support for meta-emotions along such lines also comes from appraisal theories. Such theories emphasize that emotions are produced by appraisals of how a situation is relevant for the subject's psychic constitution and how it promotes his goals and preferences. For example, Frijda (1986) has famously suggested that emotions are psychic "relevance detectors". In a recent paper, leading appraisal theorists summarize the relevant notion of appraisal as "a process that detects and assesses the significance of the environment for well-being" (Moors et al. 2013, p. 120; cf. also

---

[6] Cf. Nussbaum (2004, p. 196), or Solomon's (1976) defense of such a strong form of cognitivism and his summary in (2003). For a recent overview with a particular focus on Solomon's work see Deigh (2013).

[7] Cf., e.g., Ben Zee'ev (2000), Ben Ze'ev (2010), pp. 47–48. Typically, multi-component views also include physiological and motor expression components. Other authors who endorse various forms of multi-component views include Goldie (2000), Roberts (2003), Pugmire (2005), Deonna and Teroni (2012), Mulligan and Scherer (2012).

Frijda 2007; Lazarus 1991; Mulligan and Scherer 2012).[8] For several decades such accounts have played a leading role in emotion research. And they comport well with our claims about meta-emotions, except that talk of "environment", etc., should be (re)construed as including intrapsychic conditions. For surely, not only are certain external situations and events of utmost concern to us, but so are our emotions. Emotions clearly have goal-conducive or -obstructive effects and—arguably more than any other mental phenomenon—they determine our psychic well-being. Although to our knowledge appraisal theories have not examined this consequence, their claims thus neatly predict that engaging with one's first-order emotions will often produce meta-emotions.[9]

Despite such considerations, it may be argued that an investigation of meta-emotions should not rely exclusively on a priori considerations. In the next section, we present a pilot study suggesting that meta-emotions manifest themselves in facial expressions. Before turning to this study, however, let us add some details and conceptual clarifications.

Influential appraisal theorists, such as Lazarus, seem to endorse a "cognitive-cause theory" of the role of appraisals in emotion theory (see Prinz 2004, p. 17). However, when a first-order emotion has *caused* some other emotion, the latter may but need not be intentionally directed at the former. Your fear of the dog may cause, or causally contribute to, your becoming angry even when this anger is not directed at this fear. (Shortly after the dog has attacked you, you find yourself in a state of profound annoyance, yet what you are annoyed about is the fact that someone just stole the park bench you had been heading for.)

Second, the term "meta-emotion", or a similar label, occasionally appears in psychological studies, e.g., of affect regulation or emotion socialization, but in such contexts it is typically used in a very broad sense. For example, a recent study of families with depressed adolescents discusses people's "meta-emotion philosophy", thereby referring to "an organized set of reactions, thoughts, and feelings toward emotions" (Hunter et al. 2011, p. 430). Other studies define a "meta-emotion structure" as "an organized and structured set of emotions and cognitions about ... both one's own emotions and the emotions of others" (Gottman et al. 1997, p. 7).[10] But not every reaction to, thought about, or cognition of, an emotion is itself an emotion, hence it's misleading to call just any such phenomenon a "meta-emotion". Here we reserve the term exclusively for higher-order *affective* states or episodes. Moreover, although it is an interesting question how a subject's emotions about other people's emotions impact the subject's own emotions, and *vice versa*, interpersonal feedback effects are not our topic. My joy at your joy might affect feelings of joy I have

---

[8] Ben-Ze'ev maintains that "[e]motions occur when a change is appraised as relevant to our personal concerns" (2000, p. 18). Even William James, who famously construes emotions as "feelings of bodily changes that follow directly the perception of the exciting fact" (1890, vol. II, p. 449) argues that what produces the emotion is "the overriding idea of the significance of the event" (James 1894, p. 518).

[9] For a more comprehensive argument concerning this point about appraisal theories see Jäger and Bartsch (2006).

[10] Media psychologists have also investigated the seemingly paradoxical enjoyment of films that elicit emotions with primarily negative hedonic valence, such as fright, horror, or disgust (Oliver 1993; Bartsch et al. 2010).

about my own joy. But meta-emotions in our sense are exclusively intrasubjective phenomena.

For present purposes we need not develop a comprehensive typology of affects and emotions. It is clear, however, that if moods or mood-like states (such as object-less melancholy, free-floating depression, euphoria, etc.) lack specific intentional objects, they cannot occur as higher-order components of a meta-emotion. The same holds for other non-intentional affective phenomena such as basic forms of affective physiological arousal. On the other hand, we propose to be terminologically liberal and to classify an emotion as a meta-emotion even if its lower-level affective object is non-intentional. So, while a "meta-mood" is a conceptual impossibility on our account, we will allow for moods to serve as the intentional objects of meta-emotions.

Talk of an emotion's "having objects" can refer to different phenomena, and some distinctions are in order. Sometimes affect ascriptions suggest that an emotion is directed at a particular ("Peter is afraid of Fido"); in other cases the object or content is propositional ("Peter fears that Fido might attack him" or "Peter fears that some dog might attack him", etc.). In the first case we may say that what the emotion is directed toward is its material object or its target (see Mulligan 2007; Brady 2013, p. 26; De Sousa 2011, p. 30).[11] Sometimes what the subject takes to be the target, or some aspect of it, is among the emotion's causes, but sometimes it is not—as is most clearly the case when the supposed target does not even exist. Emotions can have illusory objects. (Macbeth is terrified of an imaginary dagger.) An emotional object may also be called "illusory" when the emotion is elicited by ascribing properties to a target which does exist but fails to have these properties. Emotions directed toward particulars typically occur in virtue of certain features of their objects: Your fear of Fido occurs in virtue of your taking the dog to be behaving, or to be disposed to behave, aggressively toward you. Some authors call these features the emotional "focus (of attention)" (De Sousa 1987, p. 116; 2011, p. 30).

Note that in some such cases it may be argued that the *de re* ascription, which in the above example has the form "S is afraid of X", is reducible to a *de dicto* ascription ("S is afraid that p"), yet in other cases the prospects for such a reduction are less rosy. If Peter's love for Mary were reducible to propositional attitudes concerned with the fact that Mary is kind, charming, beautiful, intelligent, and so on, then he should have no objections to having her replaced by another person who shares all of Mary's adorable features. However, most of us would not want our beloveds to be replaced by qualitative clones and would regard such a replacement as a loss. Peter may have many positive propositional attitudes about Mary, which may affect his feelings in many important ways. Still, he loves *her*.

Finally, consider the "formal objects" of emotions, which demarcate the boundary conditions for an affective phenomenon to be counted as an emotion of a particular type. Thus Kenny, who introduced the notion into the contemporary debate, argues that reference to formal objects can explain the fact that "each of the emotions is appropriate ... only to certain restricted objects" (Kenny 1963, p. 192,

---

[11] Teroni (2007) calls such objects "particular objects" of emotions. Among the most detailed discussions of different kinds of object-relatedness of emotions is De Sousa's (1987), chapters 4 and 5.

p. 189).[12] For example, the formal object of envy is another person's good; the formal object of fear is expected future harm or disvalue; the formal object of sadness is significant loss. That's why one cannot envy oneself—at least not one's present self—for some good one enjoys, or be afraid of something one already knows to have happened. Moreover, formal objects define the standards for an emotion's aptness. If you fear something that does not in fact threaten to inflict some future evil or disvalue upon you, your fear is in-apt. In a weaker sense it is in-apt or inappropriate only if you have no good (subjective) reasons for thinking that some future disvalue may ensue. If your emotion *is* based on good reasons it may in some sense be regarded as appropriate (apt, adequate) even when in fact your situational appraisal is inaccurate.[13]

These observations raise a number of intriguing questions about the intentionality of emotions. However, none of them raises problems distinctive of meta-emotions. To begin with, there are the general tasks of describing an emotion's intentionality more precisely and of locating its source. The account given will depend on which general theory of the emotions one wishes to adopt. For example, reductive judgmentalists like Solomon will hold that, since they take emotions to be identical with judgments or beliefs, affective intentionality is to be analyzed in terms of the intentionality pertaining to judgments and beliefs. The perceptual-appraisal theory suggested by Prinz (2004), by contrast, analyzes the aboutness of emotions in terms of a Lazarus-inspired account of their "core relational themes" in tandem with a Dretske-style theory of mental representation. Mulligan (2007) sympathizes with a multi-component theory according to which emotions are "intentional episodes that involve seeing, hearing, feeling, remembering, expecting, judging, and so forth", and argues that affective intentionality is to be explained in terms of the intentionality pertaining to these other mental episodes and states. In any case, whatever the virtues and vices of these approaches, many questions about intentionality regarding corresponding accounts of meta-emotions will not be specific to the latter but be inherited from the general theory of emotions one has adopted. Some other features of meta-emotions, by contrast, are due specifically to their higher-order nature. For example, if the meta-emotion's affective intentional object is itself an intentional emotion, the meta-emotion is in some sense also intentionally directed at the first-order emotion's object. Let us say in such cases that the meta-emotion's primary intentional object is the first-order emotion, while its secondary object is the object toward which the first-order emotion is directed. In some such cases the quality of the higher-order act carries over to the secondary intentional relation; in others it does not. For example, when you enjoy being thrilled by a bullfight (the fight is the intentional object of your first-order emotional thrill), you may, *in virtue of your meta-emotion* of enjoying the thrill, also enjoy the bullfight. However, if you are

---

[12] De Sousa (1987) even argues that "there are as many formal objects as there are different emotion types" (p. 123), and that "formal objects do not merely constrain the emotion, they define it" (p. 126). For a critical discussion of this claim and further details about the formal objects of emotions see also Teroni (2007).

[13] For further explorations of the relation between formal objects of intentional attitudes and the latter's correctness conditions see Mulligan (2007).

ashamed of being thrilled by the bullfight, you are not thereby ashamed of the bull-fight.

Like ordinary first-order emotions, meta-emotions may have propositional or non-propositional objects. One could interpret the above quote from Henry James as telling us that Newman is, in a sense, pleased *that* he feels melancholic. However, another (perhaps more natural) reading is that his positive higher-order emotion relates in some non-propositional way to his negative first-order emotion. If this target, Newman's melancholy, actually exists, then he may be said to enjoy some higher-order *de re* pleasure toward it. Again, how exactly this is to be spelled out depends on how one construes *de re* relations. We will not pursue this question here. Yet, a promising line, we suggest, could draw on the widespread (e.g., Davidsonian) view that mental states and episodes—and hence emotions as well—constitute events. Events are particulars. So if *de re* relations require particulars as their relata, an account along such lines provides a smooth way of explaining how emotions can serve as *de re* objects of meta-emotions.

A given meta-emotion's formal object is fixed by the type of emotion the meta-emotion instantiates. Accordingly, not every emotion can occur as a meta-emotion. For example, since envy cannot occur as a reflexive emotion—it cannot be directed at oneself or one's own properties—there cannot be meta-emotional envy. You can-not envy yourself, at least not your present self, for having a certain desirable first-order emotion or emotional disposition. (Perhaps one can envy oneself, or a former state of one's self, for emotions or emotional dispositions one used to have but has no longer.) Since fear is future-directed, you cannot be meta-emotionally afraid of an emotion from which you may have suffered in the past but of which you now know that it will no longer occur. (You may have experienced fear of ghosts in your childhood, but now know that such fear will never gain sway over you again.) Moreover, as in the case of first-order fear, a meta-emotional fear is apt (appropri-ate, correct) in the stronger sense only if its emotional object really does threaten the subject with some future disvalue, and in the weaker sense only if the subject reasonably believes the emotional object to do so. While it would often be inappro-priate to be afraid of being happy, it is not inappropriate to be afraid of falling into depression.

When a first-order emotion is illusory because it misidentifies its object or ascribes properties to the object which the latter does not in fact have, the state or attitude will still be an emotion. However, suppose you feel shame about some fear you think you have, but what you take to be fear is in fact fatigue. In such a case you are not in a meta-emotional state since the first-order state is not an emotion.

Are there also meta-emotions that are consciously and explicitly directed toward a *lack* of first-order emotions? Sometimes we feel guilty about (ashamed of, proud of, etc.) having emotionally tuned out; we also have emotions about failing to have, or having rid ourselves of, certain emotions. One may also, *vice versa*, fail to have emotions about having certain emotions, or fail to have emotions—including emo-tions it would be appropriate to have—about *not* having certain emotions. In Camus's *The Stranger* one of the protagonist's most irritating features is that he has not only emotionally shut down but that he doesn't even regret doing so. Such negative cases are interesting and important, and fall within the broader territory of our topic. It

seems terminologically preferable, however, to reserve the term "meta-emotion" for cases in which, both at the lower level and at the meta-level, the subject is or was engaged in some affective state. (In Sect. 4 we will return to this topic and argue that nonetheless the absence of negative meta-emotions itself constitutes a moderate form of self-harmony or "affective self-endorsement".)

In light of the Camus example, it is worth mentioning that emotion predicates can ascribe occurrent emotions as well as affective tendencies or dispositions, which in some cases may also constitute, or develop into, full-fledged character traits. These distinctions apply to meta-emotions as well. For example, we may occurrently experience joy, happiness, annoyance, disappointment, etc., regarding affective states we currently experience; or we may enjoy or dislike detecting in ourselves some affective disposition or character trait. Likewise at the meta-level. In addition to occurrent higher-order emotional experiences we may also have long-term higher-order affective dispositions which affectively charge us with respect to certain emotions or emotional dispositions at a lower level. Meta-emotions may thus stand in synchronic or in diachronic relations to their first-order affective objects. S's meta-emotion is an intrasubjective intentional emotional state or episode that is synchronically or diachronically directed at other affective states or episodes had by S.

Note that nothing of what has been said so far commits us to the view that emotions are luminous or "transparent", in the sense that whenever a person has an emotion, she is aware of this fact or enjoys some privileged access to what kind of emotion it is.[14] And we are in any case not claiming that every emotion, in whichever circumstance it occurs, elicits some meta-emotion. The present account thus need not worry about infinite hierarchies. Nevertheless, we do not wish to deny that people experience higher-order affections beyond the second level. For example, you may feel ashamed of enjoying the sentimental feelings elicited by a cheap tear-jerker. There may also be psychologically more complex cases, as portrayed in the following passage from Kingsley Amis: A man has done something (he regards as) morally reprehensible. He comes home, "feeling a tremendous rakehell, and not liking myself much for it, and feeling rather a good chap for not liking myself much for it, and not liking myself at all for feeling rather a good chap".[15]

3 Meta-emotions and facial actions: a pilot study

Are there ways of empirically studying people's meta-emotional appraisals of their first-order emotions? In this section we present a pilot study offering preliminary evidence that certain facial expressions may indicate the presence of meta-emotions. An influential objective measurement technique for non-linguistic facial behavior is

---

[14] A more detailed argument for this claim, including analyses of various kinds of privileged access claims about affective states and episodes, can be found in Jäger (2009).

[15] Cf. Amis (1955), chapter 7. This passage has made it into Michael Clark's list of famous paradoxes (Clark 2012, pp. 183–84). Whether or not it is a paradox, the story involves meta-emotions.

Ekman and Friesen's Darwin-inspired facial action coding system (FACS).[16] Although not uncontroversial,[17] this method has been used in recent decades in a great number of psychological studies of emotions, especially in developmental, social, and clinical psychology. The examples we used in our study come from clinical psychology. We examined (footage of) two series of psychotherapy sessions and one diagnostic interview in which clients discuss, among other things, their negative first-order emotions (such as fear, sadness, "feelings of helplessness", etc.).[18] When doing so, they regularly displayed facial behavior which according to FACS and corresponding tables of emotion prediction indicated the presence of negative emotions. However, often these emotions did not conform with the emotions the clients were currently discussing with the therapist. Instead, the clients' facial displays typically suggested disapproval or "contempt". A plausible interpretation, we suggest, is that in many of these situations the clients' facial expressions indicated meta-emotional disapproval of their negative first-order emotions. (In Sect. 3.5 we discuss alternative interpretations.)

3.1 Method

FACS categorizes all visible distinguishable facial movements and assigns to each such movement a so-called Action Unit (AU). It differentiates 24 AUs. Table 1 lists the most important ones, including those figuring in our study. FACS also provides 19 codes for more coarsely defined AUs, the so-called Action Descriptors (ADs, such as "tongue show" or "lip bite"), and for gaze behavior and head positions (Ekman and Friesen 1978; Ekman et al. 2002). Emotion prediction tables relate AU-combinations to basic emotions, such as anger, sadness, fear, disgust, happiness, surprise, or contempt.[19]

As far as our argument goes, nothing commits us to a particular view about basic emotions or to the related universality theses of the Darwin-Ekman-Friesen tradition according to which facial expressions are cross-culturally invariant products of hardwired, innate affect programs. Nor do we claim that there is a simple "one-to-one" emotion-expression linkage such that facial expressions are always completely reliable indicators of inner affective states or episodes. Ekman and Friesen themselves (1975) provided empirical evidence for culturally learned display rules that guide people in whether to display emotions in a given social situation. In general, felt emotions can be masked and, *vice versa*, facial displays that sometimes indicate a particular affective

---

[16]   Another, similar coding system had first been proposed by Hjortsjö (1969).

[17]   For critical discussions see for example Russell (1995), Russell and Fernández-Dols (1997), Fernández-Dols and Ruiz-Belda (1995, 1997), Fridlund (1997), Parkinson, Fischer and Manstead (2005), ch. 6. Proponents of facial expression approaches include, in addition to Darwin (1872) and Ekman and Friesen, Izard (1971, 1994), Scherer and Wallbott (1994), Frijda and Tcherkassof (1997), Scherer and Grandjean (2008). For a helpful overview of the literature both critical and in defense of facial action studies see also Niedenthal and Krauth-Gruber (2006), chapter 4. We discuss some of the criticisms below.

[18]   The footage is archived at the Institute of Psychology, University of Innsbruck, Prof. Eva Bänninger-Huber, and can be inspected on demand.

[19]   Different theorists classify different emotions as basic. For an overview see Ekman (1982, 1999a, b). The assignments of AUs to emotions are based on a rich empirical database (see, e.g., Ekman and Rosenberg 2005).

Table 1 The FACS Actions Units (AUs) occurring in our data. For selected AUs it is also possible to score the intensity of the facial movement on five levels (from trace to maximum = A–E). Unilateral occurrence is marked either with L (left side) or R (right side)

AU number	FACS name	Muscular basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae; Depressor Supercillii; Corrugator
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
9	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
12	Lip Corner Puller	Zygomaticus Major
14	Dimpler	Buccinator
15	Lip Corner Depressor	Depressor Labii
17	Chin Raiser	Mentalis
20	Lip Stretcher	Risorius
23	Lip Tightner	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
29	Jaw Thrust	No single muscle involved, more grossly defined movement

state can occur without a corresponding emotion. However, in order for our study to be relevant we must (and do) accept the moderate hypothesis that there is a non-random relation between facial displays and emotions and, more specifically, between the kind of facial behavior the Darwin-Ekman-Friesen tradition identifies with the expression of contempt or disapproval and the corresponding emotion. *In the situations we examined*, we suggest, it is plausible to interpret at least many of the facial expressions we selected as being indicative of (meta-)emotions of disapproval. In Sect. 3.5 we discuss some more general criticisms of facial action studies and alternative interpretations of our data.

In a first step, we produced transcripts of all of the potentially relevant episodes in the therapy or interview sessions. We isolated the situations in which the clients verbally identify some current or past first-order emotions, saying things such as: "... all those fears I have ...," "I was struck by a strong feeling of sadness," etc. In order to identify the relevant first-order emotions, we also took account of the situational context. After selecting the relevant sequences, we used FACS to code the subjects' facial expressions by assigning AUs. (We employed so-called apex coding: the AUs were coded in those moments in which the facial displays reached their maximum intensity.) We then used standard tables of emotion prediction to correlate the AU-combinations with emotions (cf. Ekman and Friesen 1982; Matsumoto et al. 1991). Next we discarded all sequences where the FACS-indicated emotion did not clearly differ from the first-level emotion identified in the subject's verbal reports. For example, if the client discussed her first-order sadness with the therapist and displayed facial behaviors associated with

sadness, these could not be interpreted as evidence for the presence of a meta-emotion of sadness, since the AU combinations might have been indicative of (a revival of) the client's first-order sadness.

The pictures provided below are illustrations of the facial expressions we selected. To preserve the clients' anonymity we are not using screenshots from the original footage. Instead, we provided co-workers who are experienced FACS-coders with the AU-codes and with pictures from the footage. On this basis they mimicked the facial expressions we selected. Note that the pictures below thus do not depict facial behaviors of the controversial category of "posed expressions" but spontaneous facial behavior. In any case, the pictures are not part of the data, but merely serve as illustrations. We present them here to provide readers unfamiliar with FACS with an intuitive inroad to our AU-assignments, and to allow those who are familiar with FACS to verify them.

The clients were native Germans. The psychotherapy consisted of 20 sessions and was part of a research project that studied affect regulation processes in panic patients. The clinical interviews come from a second research project, consisting of 120 video-taped OPD (Operationalized Psychodynamic Diagnostic) interviews.[20] For the following summary, we have, under the auspices and with the help of an (American) English native speaker who is fluent in German as well, carefully translated the clients' utterances from German into English. Since we are dealing with spontaneous spoken language, the subjects' utterances are often incomplete and grammatically dubious. In order to facilitate comprehension, we have in our transcriptions occasionally inserted square brackets with supplementations.

3.2 Ms. A's contempt for her fear

Ms. A was 33 years old when severe relationship problems led her to seek therapeutic help. Her partner had left her, without explanation, for another woman. Ms. A was particularly disappointed and sad because she and her partner had just built a new house. As a result, the client experienced panic symptoms as well as severe feelings of helplessness. The psychotherapy, conducted by an experienced psychoanalytic thera-pist in his early fifties, included 20 sessions; every fifth session was videotaped, with the client's approval. Our examples include sequences from the first, fifth, and 20th therapy sessions.

In the first session the most important topics are Ms. A's feelings of sadness and helplessness. She further reports intense anxiety when staying alone in her (and her former partner's) house. While talking about this, she appears very depressed. She talks with a quiet voice and frequently interrupts her speech with long pauses. About 42 min after the beginning of the session the therapist asks Ms. A how she had come to notice her anxiety. With a moan she answers "Well ...", and after a short interruption continues with: "... It was just every noise, and everything that struck me as weird,

---

[20] Such interactions are a specific form of diagnostic interview which, on the basis of the patient's symp-toms, aim to assign specific mental disorders to the patient (such as depression, personality disorders, or eating disorders).

and everything. There had been lots of break-ins around and I kind of heard about it. And the neighbors too, they had a burglar alarm, and so the neighbors were robbed while they were lying in their bedrooms, they were sleeping. And I was sitting in my 750-square-foot house, and then I just found myself going out in the evenings." Both before and after this utterance she produces a unilateral AU 14 (picture 1, time markers 42:21:24, and 42:23:62). In terms of emotion prediction, AU 14—both unilateral and bilateral—is generally regarded as an unambiguous indicator of contempt (aversion, disapproval, deprecation).[21] Our hypothesis is that Ms. A disapproves of the fear she experienced while alone in the house. The same pattern of facial expression and verbally identifying fear reoccurs about thirty seconds later, which lends further credence to this interpretation: Ms. A once more talks about the fear she experienced while being alone in the house: "I was sitting in the house and ...". She interrupts her sentence, shakes her head, and produces a unilateral AU 14.

For reasons of space we omit one situation with similar results and jump to the next sequence. It starts about 23 min after the beginning of the fifth session. The period immediately before is characterized by very long interruptions in the client's speech, and she appears very depressed and passive. The therapist tries in a supportive way to unravel the reasons for Ms. A's mood. After an 18-second pause the client introduces a new topic and reports: "On Sunday, no, on Monday, I was scared for the first time again after a long while". The word "first" is accompanied by AU 10B (time marker 23:8:50, picture 2), an indicator of contempt. Immediately after finishing this sentence she shows AU-combination 4 + 7 + 29, followed by 4 + 7 + 10B + 14 + 23 (pictures 3 and 4). This AU-combination is, according to FACS, a very intense negative expression, indicating a blend of anger, disgust, and contempt. Additionally, the patient shakes her head slightly and touches her throat, which is taken to indicate tension. Once more, these behaviors may express the client's strong negative higher-order emotional appraisal of her fear. (Her disapproval may have been enhanced by the fact that she experienced her fear as a relapse.)

The third and the fourth sequence we selected come from the 20th and last therapy session. The therapist opens the session by asking "Well?," thereby referring to the client's reaction to the videotape from the first session which he had just shown her. The client begins to describe her impressions and compares her emotions from the beginning of the therapy with the emotions she is experiencing now. "I can't defend myself," she reports about her earlier feelings. "This was a very central point for me, and I think what drove me so crazy at the time ... apart from the pain of the break-up and from …" At this point she shows an intense unilateral AU14 (1:46:46, picture 5), an indicator of contempt. She continues by saying: " ... the hurt feelings ... but what was really kind of important to me at the time, what's still important and somehow is important in every relationship, is that I get to this point where I have the feeling that I've been treated unfairly, or that I can't defend myself, so that I then find myself in this mixture of sadness, hurt feelings, and rage." Here Ms. A's mimic behavior suggests a meta-emotion which reflects her disapproval of her feelings of helplessness, her sadness, and her hurt feelings that turned into a form of rage she

---

[21] See, e.g., Ricci Bitti et al. (1989), Matsumoto (1992). "Contempt" is the standard term FACS studies and theories of "basic emotions" use here. It typically serves as an umbrella term for emotions of disapproval.

was unable to convert into action. Just before she says "sadness", the AU-combination 10B + 17, once more an expression of contempt or disapproval, appears on Ms. A's face (02:10:02, picture 6). Additionally, she slightly shakes her head. While uttering "sadness" she narrows her eyes by AU 7 (02:12:28, picture 7), which according to FACS is an element of the prototypical expression of anger. Both expressions may plausibly be interpreted as expressions of meta-emotions stemming from the patient's disapproval of her sadness and hurt feelings as experienced at the time she started her psychotherapy.

picture 1   picture 2   picture 3   picture 4

picture 5   picture 6   picture 7

3.3 Policemen don't weep!

Mr. P was 29 years old and had received the diagnosis "major depression". During the OPD-interview he was in stationary clinical treatment. The interviewer was an experienced psychiatrist and psychoanalyst in his early fifties. Apart from the classical symptoms of depression (feelings of worthlessness, loss of energy, etc.), Mr. P suffered from several severe psychosomatic symptoms such as amblyopia and nausea or stomach problems, as well as from drug and alcohol addiction. Mr. P experienced these symptoms as incompatible with his role as a police officer, and they impeded him from exercising his job properly. Furthermore, he suffered from problems in relationships.

As a child, he had been negatively affected by conflicts between his parents. When they were about to separate, his mother had asked him to decide whether to stay with her or with his father. However, the latter was not an option for Mr. P, so he tells the interviewer. In such moments he had felt pressured, and usually withdrew to his room. At this point (time marker 52:47:10) AU-combination L14 + L20 + B28 appears on his face (picture 8), and he says: "I never asked myself this question [whether I should

live with my mother or with my father] at all". He pauses, licks his lips, and continues: "In such moments I always was very disturbed; I used to cry a lot in my room." While vocalizing the word "cry" he rubs his eyes and raises his eyebrows (AUs 1 + 2, picture 9). According to FACS theory, this AU-combination is a so-called illustrator that emphasizes the emotional significance of his statement. "Always, whenever these fights [between my parents] happened, I was pretty down", Mr. P continues. At the end of this sentence, he manifests AU 14, an indicator of contempt or disapproval (picture 10). Again, the client's expression of disapproval may plausibly be interpreted as resulting from his negative evaluation of the grief and sadness he reports. This seems to hold both for his reported feelings from his childhood as well as for his apparent present feelings of sadness.

picture 8   picture 9   picture 10
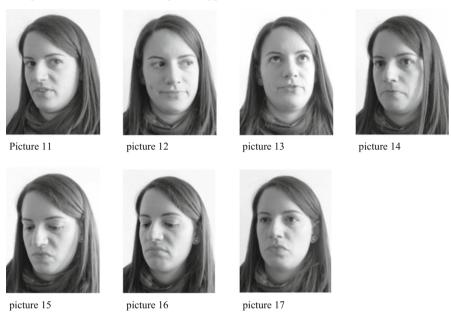
3.4 Disliking depression

Ms. C was 26 years old at the time of the interview and in stationary treatment. Like Mr. P, her diagnosis was "major depression". She was interviewed by the same interviewer as Mr. P. At the beginning of the sequence, the interviewer asks her to describe her most important negative emotional experiences. She responds by talking about her relationship problems, especially her fear of being abandoned. When she says "... fear of being abandoned", AU10B occurs in her face (07:23:21, picture 11), immediately followed by AU 14 (07:26:14, picture 12). Both expressions are indicators of contempt or deprecation. She adds the phrase "[I have] worries about the future" (07:28:09, R10B, picture 13), and once more an expression of contempt appears (07:28:19, AU1 + 2 + 15B + 23, picture 14). In the contexts at hand, these facial patterns are naturally regarded as expressions of the client's negative emotional appraisals of her intense fears and worries.

   In the next micro-sequence the patient talks about her feelings of hopelessness and depression. Remembering her grandmother's death, she says: "I really did experience this as a ....." She interrupts herself and shows a very intense expression of disapproval or contempt (33:46:24, L10C + R10E + 15B + 17, picture 15). She finishes the sentence by saying: "... very sad event". The word "very" is accompanied by the AU-combination 10B + 17, again an indicator of contempt (33:47:18, picture 16). Here, the contempt may be viewed as an affective evaluation of her feelings of sadness. "The

depressed moods", she continues, "are very difficult for me to bear when I am feeling lethargic; this totally empty feeling of being swallowed from all directions, this is the worst for me." The word "empty" is accompanied by AU-combination 7+R10B+15B (45:04:20, picture 17), a blend of an anger-contempt expression. Again, this may be an expression of the client's negative appraisal of her first-level emotion "sadness".

Picture 11   picture 12   picture 13   picture 14

picture 15   picture 16   picture 17

3.5 Discussion

These observations, we suggest, can plausibly be interpreted as providing preliminary empirical evidence that meta-emotions sometimes manifest themselves in facial behavior. However, alternative explanations of our data must be considered. Moreover, FACS, although widely employed, is not uncontroversial, and some critics may have more general reservations about applying facial-action analyses in this context in the first place.[22] In particular, it has been argued that facial displays often function as strategic social signals, or that so-called audience effects have significant impact on facial behavior which should therefore not be regarded as spontaneous expressions of experienced emotions. Before turning to alternative interpretations of our data, a general comment on such reservations is in order.

We do not wish to deny the role of audience effects on intentional and spontaneous facial behavior, or the "social situatedness" of at least many emotions and much affective behavior in many circumstances. The question is whether, in situations where these effects are operative, this indicates that the emotions corresponding to the facial displays are not present or not genuinely experienced. In any case,

---

22 An anonymous referee has pressed us on this point.

what matters for our purposes is whether in the situations we examined the subjects' facial expressions can plausibly be interpreted as indicating meta-emotional appraisals. We think they can. For example, Griffiths and Scarantino have recently argued for a "situated-perspective account" of emotion which, in contrast to both cognitivist and Neo-Jamesian approaches, emphasizes that emotional episodes involve "an ongoing exchange of emotional signals (e.g., facial actions, tones of voice)" between communication partners (2009, p. 438). However, when the authors maintain that emotional signals "can be produced strategically without becoming mere pretences of emotion" (p. 439), they seem to concede that such signals can also *express* emotions. Social signals have contents, and surely the content of a facial signal can precisely consist in the fact that the subject has a certain emotion. It is often socially useful or appropriate to use your facial expressions to let someone in on what you are in fact feeling—say in empathizing with a friend who has experienced a tragedy. In short, social-signal approaches and expression accounts need not be antagonistic.[23] The idea that, as Fridlund and Chovil once put it, "facial displays … are more likely to be emitted when there is a potential recipient, when they are useful in conveying the particular information, and when that information is pertinent or appropriate to the social interaction" (Chovil and Fridlund 1991, p. 163) is perfectly in line with our view, and it fits our studies well. After all, the clients were talking to someone. And a plausible interpretation of their facial behaviors is that they indicate meta-emotional appraisals of the subjects' first-order emotions. Further, it is reasonable to think that we are often more likely to experience meta-emotions in the presence of another person than we are by ourselves. The reason is that we will be aware of their perspective on us. Being aware of someone's perspective on us can influence us to regard ourselves from a more self-reflective distance, inducing us to consider our first-order emotions and to experience affective responses to them.

Despite all this there are alternative interpretations of the data, in which audience effects might play an important role. For example, there is evidence that people sometimes engage in facial mimicry, which may raise the question whether emotion prediction by FACS is appropriate at all for measuring emotions the subject has before interacting with a communication partner.[24] Our response is that, in the situations we studied, the subjects' interaction partners were experienced professional psychotherapists whose facial expressions remained relatively neutral in the conversations. The footage was produced in a split-windows format that constantly displayed the interviewer as well. This allowed us to verify that in the sequences we chose the clients' facial expressions did not mimic those of the psychotherapist or the interviewer.

However, assuming that the clients' expressions really are indicative of their own emotions, another question is whether these really were directed at present or past first-order emotions themselves or rather at the fact that these were displayed in situations

---

[23] Parkinson, Fischer and Manstead, in a detailed chapter on "Moving Faces in Interpersonal Life", argue that "rather than claiming that facial movements display social motives *instead* of expressing emotions, … it is possible to conclude that they communicate both kinds of information in different circumstances" (2005, p. 169, our emphasis). We believe that they can do both even in one and the same situation.

[24] For a detailed discussion see, e.g., Niedenthal and Krauth-Gruber (2006), chapter 4, or Parkinson et al. (2005), pp. 170–174.

the subjects deem inappropriate. We do not wish to deny that there are social "feeling rules" concerning the propriety of outward emotional *expression* (see Hochschild 1979). However, in our cases it seems unlikely that the clients' disapproval was directed at the fact that they violated such feeling rules. For example, when Ms. A talks about her fears she explicitly considers situations in which she was on her own and hence when no social expectations were immediately operative. Analogous observations apply to Mr. P who reports to have been crying when he was alone in his room.

Still, might not the facial displays function as social signals to the therapist about the client's awareness of what she perceives to be the (mutually agreed upon) regrettability of their emotional condition?[25] This cannot be ruled out. However, once more the question is whether, or under what interpretation, this hypothesis would be incompatible with the suggestion that a meta-emotion is expressed. Both the thesis that the client ascribes sympathy to the therapist and that she regards her situation to be generally lamentable according to certain social standards are compatible with the hypothesis that the client herself really does lament her situation. It doesn't follow that it's all fake.

Another question is whether Ms. A might not be expressing contempt not at her first-order emotions but at her ex-partner for having abandoned her in a difficult situation. Similarly, mightn't the second subject for example be expressing contempt at his parents for their miserable behavior in the past, rather than at his own emotions? As it stands, this does not appear likely. Note that the fact that Ms. A's partner has left her, plus the personal and social effects this had on Ms. A, is the constant dominating topic of all of her therapy sessions which we examined. Yet only at certain points—namely when she talks about her experiences of fear, sadness, disappointment, etc.—does she display the expressions discussed above. But even so, there may be something to the present objection. It might be that the subjects's expressions displayed their meta-emotional appraisals of their first-order emotions *as resulting from certain actions and events*. Thus, a more comprehensive description of the intentional object of one of Ms. A's meta-emotions might be that she is contemptuous about her anxiety *as caused by her ex-partner's having left her*. Similarly with Mr. P. The objects of his disapproval may plausibly be described as his past (and present) grief and depression *as caused by his parents' behavior*. So, the upshot is *not* that the clients' purported meta-emotions are not directed at their first-order emotions. It merely shows the meta-emotions might be somewhat more precisely described as being directed at first-order emotions which (from the subject's point of view) can be partly individuated by reference to their causal histories.

Future research could pursue these questions by investigating other kinds of interaction. For example, instead of analyzing conversations from psychotherapy sessions and diagnostic interviews, everyday situations could be examined in which people evaluate their emotions. Promising examples may include certain kinds of public speech, or interviews with rescue workers, firemen, soldiers, nurses, etc., in which the subjects reflect on their first-order emotions while the normative and social contexts make it likely that they will also evaluate these emotions. Moreover, lab situations could be

---

[25] Thanks to an anonymous referee for this objection and the next.

created in which emotions are induced and measured. The subjects could subsequently be asked about how they felt, while it should be clear in the context what the appropriate or socially expected stance toward the first-order emotion is. People's facial or other physiological behavior could then be analyzed and if, in addition, self-report measures are used, potential audience effects could be minimized.

4 Meta-emotions and the self

Among the most interesting implementations of what has been said so far, we believe, is the significance of meta-emotions for hierarchical or "divisional" accounts of self-integration and psychic harmony. Perhaps the most influential theories of this stripe are Frankfurt-style accounts which argue that human persons engage in higher-order self-appraisals. Let's take our cue from Frankfurt and start with an elegant recent summary of the position he has developed over some decades. "We put considerable effort into trying to get clear about what we are really like", Frankfurt argues (2006, p. 1). We don't accept ourselves just as we come but have attitudes about our own mental states and attitudes:

> "It is our peculiar knack of separating from the immediate content and flow of our own consciousness and introducing a sort of division within our minds. This elementary inward maneuver establishes an inward-directed, monitoring oversight. … When we divide our consciousness in this way, we objectify to ourselves the ingredient items of our ongoing mental life. It is this self-objectification that is particularly distinctive of human mentality. We are unique (probably) in being able simultaneously to be engaged in whatever is going on in our conscious minds, to detach ourselves from it, and to observe it—as it were—from a distance. We are then in a position to form reflexive or higher-order responses to it" (Frankfurt 2006, p. 4; cf. also 2004, pp. 17–18).

Space does not permit discussing whether the capacity Frankfurt has in mind is distinctive of humans. Instead, our question is what exactly is required to engage in the higher-order evaluations which bring us to either accept or distance ourselves from our selves.

Frankfurt's response is his famous hierarchical theory of desires. Persons, he argues, "form desires regarding their own desires—that is, regarding both what they want to want, and what they want not to want" (2004, p. 18). Let us adopt Frankfurt's terminology, understanding a person's effective first-order desires, i.e., desires that move her all the way to action (or would do so, if unimpeded) as the person's will. If a subject has the higher-order desire that one of her lower-level desires be her will, Frankfurt calls that higher-order desire a "higher-order volition". So, according to this theory, an action may result either from a will which the agent has a higher-order volition to have, or from a will he does not want to have. If an unwilling drug addict shoots up, a first-order desire that he desires not to have, or that he desires to remain unfulfilled, has triumphed over him. By contrast, the willing addict who takes the drug experiences no such conflict. He has no second-order volition that disapproves of his will. Frankfurt has also famously spelled out this idea in terms of a person's

"wholehearted identification" with her desires. If the desires that move us to act are such that we do not want them to be effective, he argues, "we are moved to act as we do without wanting wholeheartedly to be motivated as we are" (1987, p. 163; cf. also 2004, p. 19). Wholehearted self-endorsement, or a person's identification with him- or herself, by contrast, is constituted roughly by harmony between her higher-order volitions and the first-order desires that constitute her will. The relevant first-order desires must be approved of by higher-order volitions.

The view, though debatable in many of its details, has found sympathy among many eminent philosophers.[26] And yet there is one fundamental question which, as far as we can see, has not received much systematic attention to date. We do not wish to deny that Frankfurt articulates incisive insights into the nature of psychic harmony. However, those aspects of our mental lives which are liable, at times, to produce intrapersonal disintegration appear to be much richer and more multifaceted than Frankfurt-style accounts of volitional hierarchies suggest. A person's psychic identity at a given time is determined by her complete psychic profile at that time. And, essential as desires are, there is more to this profile than a hierarchy of desires. In particular, we maintain, the part of a person's psychic identity that is accountable for her tendencies to act in certain ways is also and essentially determined by her *emotions*. To begin with, there are first-order emotions, which do not reduce to first-order desires. Then, as we have argued in this essay, there are meta-emotions, and these are importantly distinct from desires or volitions at any level.

Frankfurt does not develop a substantive theory of desiring. Apart from his analyses of volitional hierarchies, he wishes to be very liberal in his usage of the term "desire". For example, in one of his early articles pioneering a hierarchical volitional account of psychic unity, he remarks that desires do not require their subjects to have any introspectible emotional response, first-person privileged access, conscious recognition, or inter-volitional coherence (cf. 1971, p. 13). Nowhere to our knowledge does Frankfurt supplement such negative characterizations with a positive theory about the nature of desires. However, with very few exceptions there is wide agreement among philosophers (and psychologists) that emotions, while often having causal links with desires, form a logically independent, *sui generis* class of mental items. One of the most fundamental differences, for example, is their "direction of fit". While conative attitudes such as desires have a world-to-mind direction, it does not seem right to say that an emotion is "fulfilled" or "satisfied" if and only if the course of the world unfolds in a certain way. (Think, e.g., of being angry at someone, fearing something, etc.) A fuller treatment of this topic would have to draw on a detailed theory of desires, which is beyond the scope of this article. It may be stated, however, that first, if—as most theories of desires *and* theories of the emotions agree—emotions cannot be reduced

---

[26] These include, to mention just a few, Michael Bratman, Sarah Buss, Wayne Davis, John Martin Fischer, Christine Koorsgaard, Richard Moran, Eleonore Stump, Thomas Scanlon, Michael Taylor, Gary Watson, Susan Wolf, Linda Zagzebski, and many others. For some representative discussions and Frankfurt's replies, see for example the essays in Buss and Overton (eds.) (2002).

to desires,[27] the relevant first-order objects of higher-order self-appraisals include not just desires but emotions too.

Second, as we have extensively shown in this essay, often the kind of higher-order appraisal in which we take a stance toward our first-order psychic life progresses via emotions. Arguably, Frankfurt himself offers the conceptual framework for an argument for this conclusion. According to Frankfurt, wholeheartedness with respect to a given psychic phenomenon consists in the subject's identification or "full satisfaction" with that phenomenon, while such satisfaction is "constituted just by the absence of any tendency or inclination to alter its condition" (1992, p. 104). Now, it is widely agreed that one essential characteristic of emotions is their potential to produce states of "action readiness" (see, e.g., Frijda 1986, or Frijda and Tcherkassof 1997). Emotions motivate us to engage in characteristic behavior: Different emotions, in different situations, motivate us to engage in different behaviors. Yet in general, emotions which the subject experiences as positive will motivate her to uphold and maintain the situations that (she thinks) bring about these emotions, whereas emotions she perceives as negative will foster tendencies to alter the situation. Having negative emotions about one's psychic condition, therefore, indicates a lack of "full satisfaction" with it. As argued at length, however, the relevant first-order psychic states or episodes include emotions. It follows that Frankfurt's satisfaction condition for self-harmony and wholeheartedness is fulfilled only if the subject does not entertain any negative meta-emotions toward her first-order emotions.[28] While ordinary first-order emotions motivate us to either preserve or alter external situations, meta-emotions motivate us to retain or alter the relevant first-order emotions we have. If you feel happy about being joyful, you will normally have the tendency to retain and perpetuate that first-order emotion. Feeling ashamed of (guilty about, annoyed about, etc.) being angry will normally motivate you to control your anger, try to cool down, and so forth.[29]

Two kinds of situation fulfill this requirement for being satisfied with one's self. To see this, consider first *volitional* satisfaction with one's first-order desires. Frankfurt tends to focus on situations in which first-order desires comply or fail to comply with second-order desires that positively want them. However, there is also a weaker sense in which we may enjoy volitional harmony. Not only are we free from volitional disharmony when we positively approve of our first-order desires or our first-order will. Things already go relatively smoothly when we have no desire *not* to harbor the first-order desires we detect in ourselves. Similarly for emotions. In light of this observation, let us distinguish strong from moderate forms of affective self-endorsement. Let us say

---

[27] For more on this point and further arguments for distinguishing emotions from desires see, e.g., Deonna and Teroni (2012), chapters 1 and 3.

[28] Note that if you have a first-order emotion with a negative hedonic tone it need not be negative in other respects (e.g., it needn't be normatively or morally inappropriate). For example, you may be tormented by deep grief about the loss of a loved one but accept this grief as perfectly appropriate.

[29] Frankfurt occasionally describes our first-order "psychic raw material" in terms of "cognitive, affective, attitudinal, and behavioral processes" or "feelings, desires, ... attitudes and motives" (Frankfurt 1992, p. 103, 2006, pp. 5–6). Accordingly, some commentators have touched on the topic of emotions. For example, Buss (2002, p. xi) characterizes Frankfurt's view of self-alienation by saying that "most of us find it difficult to identify wholeheartedly with all of our emotions, desires, and inclinations". However, Frankfurt unperturbedly continues to portray the kind of (dis)harmony at issue in terms of *volitional* hierarchies.

that a person enjoys *moderate affective self-endorsement* if she has no negative meta-emotions about her first-order emotions. She enjoys *strong affective self-endorsement* if she has positive meta-emotions about her first-order emotions.

To be sure, these are two ends of a spectrum. Self-endorsement admits of degrees. For example, sometimes you will strongly affectively endorse some of your emotions, while accepting others only in the moderate sense just sketched. In any event, what is important is that phenomena such as "self-endorsement" or "distancing oneself from one's self", which insightful work by Frankfurt and friends tends to understand in terms of higher-order volitions, should also be spelled out in terms of higher-order emotions. More of this territory needs to be explored, especially in light of a more detailed theory of the relation between emotions and desires. But for now it is time to call it a day.

5 Conclusion

Examining the ways in which persons reflectively engage with their selves, philosophers have traditionally focused on higher-order beliefs and especially higher-order desires. However, a number of observations, we have argued, suggest that one of the most pervasive and important ways in which we appraise ourselves employs higher-order emotions. We have explored the phenomenology of various meta-emotions, discussed their intentional structure, and examined the conditions under which they occur. We then presented a pilot study suggesting that meta-emotions may manifest themselves in facial expressions and that they figure among the psychic mechanisms by which people engage with their first-order emotions. Finally, we have outlined the significance of meta-emotions for the most influential family of theories of psychic harmony. "I hate and I love", Catullus writes in a famous poem about his feelings for Lesbia. "Why do I do this, you might ask. I don't know. Yet this is what I feel, and I am tormented."[30] Perhaps Catullus has mixed first-order feelings for Lesbia. But he also feels tormented by his first-order feelings. Philosophy might not be able to help Catullus with his emotions. But it can try to provide a conceptual framework for understanding the kind of condition he grapples with.[31]

References

Amis, K. (1955). *That uncertain feeling*. London: Victor Gollancz.
Augustine. (1992). (CF): *Confessions*. J. J. O'Donnell (Ed.). Oxford: Clarendon Press.
Aquinas, T. (1901). (ST): *Summa Theologiae*. Prima Secundae (Ed.). Marietti, Rome.
Bartsch, A., Appel, M., & Storch, D. (2010). Predicting emotions and meta-emotions at the movies: The role of the need for affect in audiences' experience of horror and drama. *Communication Research*, *37*, 167–190.
Ben Zee'ev, A. (2000). *The subtlety of emotions*. Cambridge, Mass: MIT Press.
Ben Ze'ev, A. (2010). The thing called emotion. P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 41–62). Oxford: Oxford University Press.

---

[30] "Odi et amo. Quare id faciam fortasse requiris. Nescio. Sed fieri sentio et excrucior." Catullus (CAR), Carmen 85.

[31] For helpful discussions we wish to thank Katherine Dormandy and two anonymous referees for *Synthese*.

Brady, M. S. (2013). *Emotional insight—The epistemic role of emotional experience*. Oxford: Oxford University Press.

Buss, S. (2002). Introduction, in Buss and Overton (2002), pp. xi–xx.

Buss, S., & Overton, L. (Eds.). (2002). *Contours of agency: Essays on themes from Harry Frankfurt*. Cambridge, Mass, London: MIT Press.

Catullus, G. V. (1958). (CAR): *Carmina*. R. A. B. Mynors (Ed.). Oxford: Oxford University Press, 1958.

Chovil, N., & Fridlund, A. J. (1991). Why emotionality cannot equal sociality. *Journal of Nonverbal Behavior*, *15*, 163–167.

Clark, M. (2012). *Paradoxes* (3rd ed.). New York: Routledge.

Darwin, C. (1872). *The expression of the emotions in man and animals*. P. Ekman (Ed.). New York: Oxford University Press, 1998.

De Sousa, R. (1987). *The rationality of emotions*. Cambridge, Mass: MIT Press.

De Sousa, R. (2010). The mind's Bermuda Triangle: Philosophy of emotions and empirical science. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 95–117). Oxford: Oxford University Press.

De Sousa, R. (2011). *Emotional truth*. Oxford: Oxford University Press.

Deigh, John. (2013). Introduction—Robert Solomon's theory of emotions in retrospect. In J. Deigh (Ed.), *On emotions* (pp. 1–13). Oxford: Oxford University Press.

Deonna, J. A., & Teroni, F. (2012). *The emotions*. London, NY: Routledge.

Ekman, P. (1982). *Emotion in the human face* (2nd ed.). Cambridge: Cambridge University Press.

Ekman, P. (1999a). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Chichester: Wiley.

Ekman, P. (1999b). Facial expressions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 301–320). Chichester: Wiley.

Ekman, P. & Friesen, W. V. (1975). *Unmasking the face—A guide to recognizing emotions from facial expressions*. Englewood Cliffs: Prentice Hall, repr. Cambridge, Mass.: Malor Books, 2003.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman P. & Friesen, W. V. (1982). EMFACS coder's guide, (unpubl. manuscript).

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system. The manual on CD Rom*. Salt Lake City, UT: Network Information Research Corporation.

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals. Basic and applied studies of spontaneous expression using the facial coding system (FACS)* (2nd ed.). Oxford, New York: Oxford University Press.

Feagin, S. (1983). The pleasures of tragedy. *American Philosophical Quarterly, 20*, 95–104. Repr. In A. Neil and A. Ridley (Eds.), *Arguing About Art, 1*. New York et al.: McGraw-Hill 1995, pp. 204–218.

Fernández-Dols, J., & Ruiz-Belda, M.-A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic games. *Journal of Personality and Social Psychology*, *69*, 1113–1119.

Fernández-Dols, J.-M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 255–274). New York: Cambridge University Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. Repr. In H. G. Frankfurt, *The importance of what we care about—Philosophical essays* (pp. 11–25). Cambridge: Cambridge University Press, 1988.

Frankfurt, H. G. (1987). Identification and Wholeheartedness. Repr. In H. G. Frankfurt, *The importance of what we care about—Philosophical essays* (pp. 159–176). Cambridge: Cambridge University Press, 1988.

Frankfurt, H. G. (1992). The faintest passion. Repr. In H. G. Frankfurt, *Necessity, volition, and love* (pp. 95–107). Cambridge: Cambridge University Press, 1999.

Frankfurt, H. G. (2004). *The reasons of love*. Princeton and Oxford: Princeton University Press.

Frankfurt, H. G. (2006). *Taking ourselves seriously and getting it right*. Stanford, Cal.: Stanford University Press.

Fridlund, A. J. (1997). The new ethology of human facial expressions. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 103–129). New York: Cambridge University Press.

Frijda, N. H. (1986). *The emotions*. New York: Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, NJ: Erlbaum.

Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). New York: Cambridge University Press.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford: Clarendon Press.

Gottman, J. M., Katz, L. F., & Hooven, C. (1997). *Meta-emotion: How families communicate emotionally*. Mahwah, NJ: Erlbaum.

Griffiths, P. E. (1997). *What emotions really are*. Chicago, London: University of Chicago Press.

Griffiths, P., & Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 437–453). Cambridge: Cambridge University Press.

Hjortsjö, C.-H. (1969). Man's face and mimic language (orig.: Människans ansikte och mimiska språket). Malmö: Studentlitteratur.

Hochschild, A. R. (1979). Emotion work, feeling rules and social structure. *American Journal of Sociology*, *85*, 551–575.

Hugo, V. (1866). *Les Travailleurs de la Mer*, E. Testard (Ed.), 1891.

Hunter, E. C., et al. (2011). How do I feel about feelings? Depressed and healthy adolescents. *Journal of Youth and Adolescence*, *49*, 428–441.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288–299.

Jäger, C. (2009). Affective ignorance. *Erkenntnis*, *71*, 123–139.

Jäger, C., & Bartsch, A. (2006). Meta-emotions. *Grazer Philosophische Studien*, *73*, 179–204.

James, H. (1876/77). *The American*. Repr. New York: Penguin Classics, 1981.

James, W. (1890). *The principles of psychology* (Vol. I and II). New York: Henry Holt and Company.

James, W. (1894). The physical basis of emotion. *Psychological Review*, *1*, 516–529.

Kenny, A. (1963). *Action, emotion and will*. London: Routledge.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion*, *16*, 363–368.

Matsumoto, D., Ekman, P., & Fridlund, A. (1991). Analyzing nonverbal behaviour. In P. W. Dowrick (Ed.), *Practical guide to using video in the behavioural sciences* (pp. 153–165). New York: Wiley.

Mendonça, D. (2013). Emotions about emotions. *Emotion Review*, *5*, 390–396.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*, 119–124.

Mulligan, K. (2007). Intentionality, knowledge and formal objects. In T. Rønnow-Rasmussen, B. Petersson, J. Joseffsson, & D. Egonsson *Hommage à Wlodek, electronic Festschrift for Wlodek Rabinowicz*. www.fil.lu.se/hommageawlodek.

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, *4*, 345–357.

Niedenthal, P. M., & Krauth-Gruber, S. (2006). *Psychology of emotion*. New York, Hove: Psychology Press.

Nussbaum, C. (2013). Emotion and personal identity. In J. Deigh (Ed.), *On Emotions* (pp. 198–214). Oxford: Oxford University Press.

Nussbaum, M. (2004). Emotions as judgments of value and importance. In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 183–199). Oxford: Oxford University Press.

Oliver, M. B. (1993). Exploring the paradox of sad films. *Human Communication Research*, *19*, 315–342.

Parkinson, B., Fischer, A. H., & Manstead, A. S. R. (2005). *Emotions in social relations*. New York: Psychology Press.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.

Pugmire, D. (2005). *Sound sentiments: Integrity in the emotions*. Oxford: Clarendon Press.

Ricci Bitti, P. E., Brighetti, G., Garotti, P. L., & Boggi-Cavallo, P. (1989). Is contempt expressed by pancultural facial movements? In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 329–339). Amsterdam: Elsevier.

Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge: Cambridge University Press.

Robinson, J. (2004). Emotion—Biological fact or social construction? In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 28–43). Oxford: Oxford University Press.

Roosevelt, F. D. (1933). Inaugural address, March 4, 1933. In S. Rosenman (Ed.), *The public papers of Franklin D. Roosevelt, Vol. 2: The year of crisis, 1933* (pp. 11–16). New York: Random House, 1938.

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin*, *118*, 379–391.

Russell, J. A., & Fernández-Dols, J. M. (1997). What does facial expression mean? In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 3–30). New York: Cambridge University Press.

Russell, J. A., & Fernández-Dols, J. M. (Eds.). (1997). *The psychology of facial expression*. Cambridge: Cambridge University Press.

Scherer, K. R., & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, *22*, 789–801.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for the universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, *66*, 310–328.

Solomon, R. (1976). *The passions—The myth and nature of human emotion*. Garden City, NJ: Anchor Press/Doubleday.

Solomon, R. (2003). Emotions, thoughts, and feelings: What is a 'cognitive theory' of the emotions and does it neglect affectivity? In A. Hatzimoysis (Ed.), *Philosophy and the emotions* (pp. 1–18). Cambridge: Cambridge University Press.

Teroni, F. (2007). Emotions and formal objects. *Dialectica*, *61*, 395–415.

Tolstoy, L. (1869). *War and Peace* (critical Ed.). Cambridge: Cambridge World Classics, 2010.