Emotions and the body: Testing the Subtraction Argument

Rodrigo Díaz, University of Bern

(Forthcoming in *Philosophical Psychology*, please cite the published version)

Abstract

Can we experience emotion without the feeling of accelerated heartbeats, perspiration, or other changes in the body? In his paper "What is an emotion", William James famously claimed that "if we fancy some strong emotion and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind" (1884, p. 193). Thus, bodily changes are essential to emotion. This is known as the Subtraction Argument. The Subtraction Argument is still used to support Embodied theories of emotion. Proponents of Disembodied theories, however, have different intuitions regarding the role of bodily changes in the experience of emotion. In order to advance this debate, I conducted a series of studies testing people's intuitions regarding the Subtraction Argument. Results show that, against James' claims, most people consider that their hypothetical (Study 1) and actual (Study 2) emotions persist in the absence of bodily feelings. Furthermore, Study 1 found that participants' responses were not related to individual differences in reflection or interoceptive awareness. These results pose a problem for Embodied theories of emotion and challenge extant explanations for the persistence of emotional experience in cases of impaired interoception.

1. Introduction

Emotions pervade our everyday life. Imagine a regular workday. In the morning, we get *angered* by the sound of the alarm clock, *enjoy* our first coffee, fear that we might be late for work, and so on. But what is anger, joy, or fear? According to one influential view, emotions are constituted by changes in one's own body, and the experience of emotion consists on feeling those changes as they occur.[1] Understood this way, anger is constituted by bodily changes such as a rise in bodily temperature, fear is constituted by bodily changes such as shivering, and so on. This view dates back to the work of William James (1884) and Carl Lange (1922), and it has been revamped in the work of contemporary authors in both philosophy (e.g. Prinz 2004) and psychology (e.g. Damasio 1994). Because of their emphasis on the role of the body, these accounts are known as Somatic or Embodied theories of emotion.

Embodied theories of emotion have some intuitive attractiveness. Ordinary language expressions and metaphors for emotions suggest that bodily changes feature centrally in our emotional experiences. For example, when we are angry, we say that our "blood is boiling". And when we are afraid, we feel "the chills". Furthermore, there is ample evidence showing that emotional experiences are accompanied by a range of potentially felt changes in cardiovascular, respiratory, and electrodermal activity (see Kreibig 2010; but also Reisenzein and Stephan 2014). However, a mere association between emotional experiences and felt bodily changes do not prove Embodied theories right. The bodily changes associated with emotional experiences could be contingent effects of the (otherwise disembodied) emotions. Thus, the crucial question is whether bodily changes are *essential* to emotion.

[1] Although some Embodied accounts identify emotions with feelings, most Embodied accounts differentiate between emotion, which is constituted by bodily changes, and the experience of emotion, which consist on feeling those bodily changes. This way, they can account for the existence of unconscious emotions (Prinz, 2004, see also Cochrane 2017).

The phenomenology of emotion, *what is like* to be in an emotional state, is supposed to show that bodily changes are not mere consequences but rather essential ingredients of emotion. Consider the following thought experiment. Imagine having an emotion (e.g. fear) and subtract from the experience of it all the feelings of bodily changes that it involves (e.g. accelerated heartbeat, sweaty palms, shivering, …). What is left? According to William James, who first presented this thought experiment, the outcome is clear. After subtracting the feelings of bodily changes, "we have nothing left behind" and "most people, when asked, say that their introspection verifies this statement" (James, 1884, p. 193). Without feelings of bodily changes, the emotional experience is not merely impoverished, but is not emotional at all. Thus, bodily changes are essential to emotion. This is known as the Subtraction Argument.

The methodology that grounds the Subtraction Argument is not free of criticisms (see, e.g. Griffiths 1997). Ideally, the claim that bodily changes are essential to emotion should be tested in cases where bodily changes are actually absent, and not just imagined to be absent. However, as James himself already noted (James 1884; James 1890), a case of "complete bodily anesthesia" is extremely rare, if not impossible (see also Laird 2007). Certain clinical conditions (e.g. spinal cord injuries) result in impaired bodily feedback, but bodily feedback is never completely absent in these cases. Because of this (see also §6), the Subtraction Argument still plays an important role in contemporary debates regarding the nature of emotion. Although using a thought experiment is not ideal, imagination might be the only way to completely subtract bodily changes from the experience of emotion.

We can find references to the Subtraction Argument in the work of Julien Deonna and Fabrice Teroni: "Emotions are psychological episodes whose nature must be understood in terms of the way the body feels. This is the conclusion James (1890/1950) draws from his famous subtraction argument" (Deonna and Teroni 2017, p. 56), or Jesse Prinz: "A number of

arguments [for emotions' embodiment] can be discerned in the pages of James and Lange. The most famous is the Subtraction Argument." (Prinz, 2004, p. 55-56).

Authors who defend Disembodied accounts of emotion, however, have made opposite claims regarding the role of bodily changes in emotional experiences. Martha Nussbaum, for example, claims that "There usually will be bodily sensations and changes of many sorts involved in grieving; but if we discovered that my blood pressure was quite low during this whole episode, or that my pulse rate never got above sixty, we would not, I think, have the slightest reason to conclude that I was not really grieving." (Nussbaum, 2001, pp. 57-58), and Peter Goldie states that "Intuitively, it might seem rather obvious that bodily feelings which are characteristic of an emotional experience are not necessary for it." (Goldie, 2000, p. 52). As James does, these authors appeal to our intuitions and first-person knowledge about the phenomenology of emotion. However, they draw the opposite conclusion. Namely, that bodily changes are *not* essential to emotion.

In order to advance the debate between Embodied and Disembodied theories of emotion, I conducted a series of experimental-philosophical studies testing people's intuitions regarding the Subtraction Argument. In the next section, I will introduce the debate between embodied and disembodied accounts of emotion. Afterwards, I will present the Subtraction Argument. The results of two new studies will show that, contrary to James' claims, most people consider that their hypothetical (Study 1) and actual (Study 2) emotional experiences persist in the absence of bodily feelings. Finally, I will discuss the results in the context of extant studies on cases of impaired interoception and extract implications for emotion theory.

Data and materials are available at https://osf.io/3u26j/.

2. Embodied or Disembodied, that is the question

Prototypical emotional experiences involve a variety of features. Most importantly, emotions seem to have both cognitive and bodily elements. Consider the following example. You are sitting at your computer and you receive a new email. The reviews of your latest paper are in. They start by mentioning that your article is not good enough for the journal. Then, they present objections to your main point that you already considered and answered in the manuscript. Finally, they tell you to look at some references that you already know about and are orthogonal to your research question. You find these comments extremely offensive and dismissive. You feel your heart pounding and your blood boiling. You clench your fist and punch the table. You are so angry.

As we see in the case above, experiencing an emotion typically involves some cognitive evaluation of the eliciting situation or event. In anger, for example, we evaluate the situation as offensive or unfair. But not only our mind is running. Our body reacts as well. In anger, our body prepares for retaliation, quickly pumping blood and clenching muscles. Paradigmatic examples of emotion involve both cognitive and bodily components. But which of them is the essence of the emotion? Can we experience emotions without one of these components?[2]

Responses to the questions above separate theories of emotion in two main camps, exemplified in the classic Lazarus-Zajonc debate (Lazarus 1984; Zajonc 1984). Lazarus gives cognition primacy, relegating bodily changes and actions to the status of mere consequences of one's cognitive evaluation of the situation. Zajonc argues that emotions can occur without any

[2] This is what Jesse Prinz calls the Problem of Parts (Prinz, 2004), the problem of determining which components are essential and which are not. Accounts that identify emotion with a combination of cognitive and bodily elements face a different problem, the Problem of Plenty, which consist of explaining how very different elements hang together to create a unitary phenomenon.

cognitive evaluation, bodily arousal being the central aspect of emotion. Lazarus' account is an example of Cognitive or Disembodied account of emotion, while Zajonc represents Non-Cognitive or Embodied theories. It is important to note that there are many different theories within each side of the Embodied vs Disembodied divide, each with its particular nuances. Reviewing all the different positions, however, is beyond the scope of this paper (for useful reviews, see Deonna and Teroni, 2012; Scarantino and de Sousa, 2018). For our purposes here, emotion theories will be classified as either Embodied or Disembodied theories.

Most Embodied theories identify emotions with perceptions of one's own bodily changes (James 1884; Prinz 2004; Laird 2007; Hufendiek 2016). But some identify emotions with embodied attitudes (Deonna and Teroni 2017). And yet others identify emotions with both bodily changes *and* representations of external objects or events (Schachter 1964; Damasio 1994; Barlassina and Newen 2014). In these latter accounts, bodily changes are not sufficient for emotion, but they are still necessary constituents. Thus, what is common to all Embodied theories is their commitment to the claim that bodily changes are essential to emotion.

Disembodied theories include theories that identify emotions with cognitive appraisals (Arnold 1960), judgments of value (Nussbaum 2001; Solomon 2003), perceptions of value (Tappolet 2016; Rossi and Tappolet 2018), or evaluative feelings (Helm 2009; Mitchell 2018). Despite these differences, however, all Disembodied theories posit that bodily changes are *not* essential to emotion. It is important to note that Disembodied theories do not deny that emotions can be accompanied by bodily changes, and thus bodily changes can feed in the experience of emotion. But, contrary to Embodied theories, Disembodied theories defend that experiencing emotion does not depend on experiencing bodily changes.

3. The Subtraction Argument

The claim that bodily changes are essential to emotion creates one of the most important divides in emotion theory. What are the arguments that support this claim? As we have seen in §1, a central line of support[3] is William James'[4] Subtraction Argument:

> "I now proceed to urge the vital point of my whole theory, which is this. If we fancy some strong emotion and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind, no 'mind-stuff out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains." (James, 1884, p. 193).

In order to defend the claim that bodily changes are essential to emotion, William James builds on a thought experiment. In this thought experiment, we are first told to imagine a strong emotional experience. Then, we are asked to subtract from consciousness all the feelings of bodily changes that it involves. The resulting intuition is that, in the absence of felt bodily

[3] Other arguments in the Embodied vs Disembodied debate are not directly about whether bodily changes are essential to emotion. Many arguments are concerned with the explanatory power of each view. For example, it is usually claimed that Embodied theories cannot account for the intentionality of emotions, the fact that emotions are directed towards objects and events in the world (e.g. fear of heights, fear that the market will crash, etc.). Conversely, Disembodied theories are sometimes accused of ignoring the phenomenology of emotion, the fact that emotions feel in certain ways. Note, however, that these challenges do not remain unanswered. For example, Embodied theories have explained the intentionality of emotion by endorsing teleosemantics (Prinz, 2004) or an action-oriented understanding of bodily changes (Deonna & Teroni, 2017; Hufendiekk, 2015). And Disembodied theories have explained the phenomenology of emotion in terms of cognitive phenomenology (Nussbaum, 2001) or perceptual phenomenology (Tappolet, 2016).

[4] In his essay "The emotions", Carl Lange makes a similar point: "Take away the bodily symptoms from a frightened individual; let his pulse beat calmly, his look be firm, his color normal, his movements quick and sure, his speech strong, his thoughts clear; and what remains of his fear?" (Lange, 1922, p. 66)

changes, the emotion disappears. Thus, as Embodied theories claim, bodily changes are essential to emotion.

The Subtraction Argument is an argument from phenomenology. It builds on our first-person knowledge of *what it is like* to experience an emotion. However, it does not rely on us enacting an actual emotional experience. The task that James proposes "is the purely speculative one of subtracting certain elements of feeling from an emotional state supposed to exist in its fulness, and saying what the residual elements are" (James, 1884, p. 193). The Subtraction Argument is thus not so different from other philosophical thought experiments such as, e.g., the Trolley dilemma (Foot 1967).[5] These two thought experiments have a similar structure: We are presented with a hypothetical case, and we are told to make a judgment about the relevant concept. In the Trolley dilemma, the judgment concerns moral permissibility. In the Subtraction Argument, the judgment is about emotion. The same way that Phillipa Foot claims that the intuitive response to the Trolley dilemma is that we should divert the runaway trolley to the tracks where it will kill fewer people, William James claims that the intuitive response to his thought experiment is that the emotional experience disappears in the absence of bodily feelings. These responses are in turn used as premises for a conclusion (Brun 2017): that the number of lives matter morally, or that bodily changes are essential to emotion.

Extensive work in moral psychology has shown that most people share Foot's intuition regarding the Trolley dilemma (Waldmann et al. 2012).[6] But do people share James' (and neo-

[5] "Be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed." (Foot, 1967, p. 2)

[6] But variations on Foot's "Switch" version of the dilemma alter the pattern of responses.

Jamesians) intuition regarding the Subtraction Argument? In the next section, I will present the results of an experimental-philosophical study designed to answer this question.

4. Study 1

In order to test whether people share James' intuitions regarding the Subtraction Argument, a simple task was designed. In this task, participants are asked to imagine a full emotional experience and then subtract the bodily feelings that it involves. I used the specific examples of subtraction that James presents in his 1884 paper. In this paper, James shows how the subtraction would work for three different emotions: grief (sadness), rage (anger), and fear. The fear case, for example, is as follows:

> What kind of an emotion of fear would be left, if the feeling neither of quickened heart-beats nor of shallow breathing, neither of trembling lips nor of weakened limbs, neither of goose-flesh nor of visceral stirrings, were present, it is quite impossible to think. (James, 1884, pp. 193-194)

The materials in this Study were modeled after this passage and the corresponding ones for sadness and anger (see footnote 8). Sampling, design, and analysis were pre-registered (https://osf.io/n2tk6/?view_only=f488b52cd4e84bafb151e2dfa1d97668). Apart from participants' intuitions regarding the Subtraction Argument, participants filled two questionnaires measuring their tendency to (1) engage in cognitive reflection and (2) attend to their own bodily states. These two traits are relevant to the task at hand, which involves *reflecting* on the role of *bodily feelings* in the experience of emotion. Thus, individual differences in reflection and bodily awareness might influence participants' responses.

4.1. Method

218 participants were recruited through Amazon Mechanical Turk and completed the survey for a monetary payment. They were randomly assigned to one out of three cases, each corresponding to a particular emotion: Fear, Anger, or Sadness. 21 participants did not complete the survey and 8 participants did not pass the control question (see below). Following the inclusion criteria established in the preregistration form, these participants were excluded. The final sample consisted of 189 participants (103 male, 85 female, 1 other, M_{age} = 36.28, *SD* = 12.39, age range 20-84). Post-hoc power analyses using G*Power showed that the study had 99% power to detect an effect of d = .5.

In the following, I will describe the procedure for the Fear case (see also Table 1). The same procedure, but different wordings, were used for sadness and anger. Two steps were used to make participants imagine a fear experience. First, participants were asked to 'Please think of a really dangerous situation and write it in the box below'. Then, they were told to 'Now imagine yourself facing that situation' and answered the question 'Would you be afraid?' on a scale from 1 ('definitely not') to 4 ('definitely yes') (from now on, I will refer to responses to this question as 'pre-subtraction emotion ratings'). As it is required that participants first imagine a full emotional experience, participants who answered 1 ('definitely not') or 2 ('probably not') were excluded from the analysis. Afterwards, participants were asked to subtract the bodily feelings from their emotional experience using James' original wording ('Now imagine that you don't feel quickened heart beats nor shallow breathing, neither trembling lips or weakened limbs, neither goose-flesh nor visceral feelings')[7] and answered the

[7] Wording for the Anger and Sadness cases were "Now imagine that you feel no ebullition in the chest, no flushing of the face, no dilatation of the nostrils, no clenching of the teeth, no impulse to vigorous action" (Anger) and "Now imagine that you don't cry or sob, you don't feel suffocation of the heart nor a pang in the breast bone." (Sadness) (see James, 1884, p. 194).

question ‘Would you still consider yourself to be afraid?’ on a scale from 1 (‘definitely not’) to 4 (‘definitely yes’) (from now on, I will refer to responses to this question as ‘post-subtraction emotion ratings’). Finally, participants were asked to justify their responses by writing a response to the question ‘Why would you still (no longer) consider yourself to be afraid?’.

Preamble	Please think of a really dangerous (sad, offensive) situation and write it in the box below. (Open text)
Pre-Subtraction Ratings	Now imagine yourself facing that situation. Would you be afraid (sad, angry)? (1 ‘definitely not’ to 4 ‘definitely yes’)
Post-Subtraction Ratings	Now imagine that you don’t feel… (James’ original wording for each emotion). Would you still consider yourself to be afraid (sad, angry)?’ (1 ‘definitely not’ to 4 ‘definitely yes’)
Justification	Why would you still (no longer) consider yourself to be afraid (sad, angry)?’ (Open text)

Table 1. Procedure for Study 1.

Apart from recording participants' intuitions regarding James' thought experiment, individual differences in cognitive reflection and interoceptive awareness were measured. Two scales were used for this: the cognitive reflection test (CRT; Frederick, 2005) and the private body consciousness scale (PBCS; Miller, Murphy, & Buss, 1981). The CRT consists of three questions. For example: 'A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much (cents) does the ball cost?'. It is considered that, in order to answer this question correctly, one has to override giving the intuitive response ($0.10). The CRT is thus supposed to measure one's tendency to engage in cognitive reflection, the tendency to respond 'on a second thought'. After the CRT scale, participants were asked to rate their familiarity with its questions on a scale from 1 (not familiar at all) to 5 (extremely familiar). The PBCS consists of 5 statements (e.g. 'I am sensitive to internal bodily tensions'). Participants rated how characteristic each statement is of them on a scale from 0 (extremely uncharacteristic) to 4 (extremely characteristic). They are given a final score, ranging from 0 to 20, which indicates their degree of awareness of their own bodily states.

4.2. Results

Exploratory analyses showed that pre-subtraction emotion ratings (*Mean* = 3.76, *SD* = .430) were significantly higher than post-subtraction emotion ratings (*Mean* = 3.09, *SD* = .843), $t(188) = 10.9$, $p < .001$, $d = .708$.[8] However, James' claim is not that emotion is significantly influenced by the absence of bodily feelings, but that emotion *disappears* in the absence of bodily feelings. Thus, the main analyses contrast participants' post-subtraction emotion ratings with James' intuitions on this respect. James' response to the subtraction question is clear: "We

[8] Non-parametric tests yielded similar results. Wilcox signed-rank test showed that pre-subtraction emotion ratings (*Median* = 4) were significantly higher than post-subtraction emotion ratings (*Median* = 3), $Z = -8.559$, $p < .001$. Furthermore, the effect was significant for all Fear, Anger and Sadness conditions (all *p*s < .001).

have nothing left behind, no 'mind-stuff' out of which the emotion can be constituted" (James, 1884, p193). Participants' responses to the post-subtraction question (*Mean* = 3.09, *SD* = .843) were significantly higher than James' answer (1 - 'definitely not'), $t(188) = 34.1$, $p < .001$, $d = 2.479$. Participants' responses to this question were also significantly higher than the scale's midpoint (2.5), $t(188) = 9.6$, $p < .001$, $d = .800$ (see Figure 1).[9] The vast majority of participants (77.3%) responded to the post-subtraction question by choosing either 'probably yes' (42.8%) or 'definitely yes' (32.4%).

Internal reliability was good for both the CRT (α = .760) and the PBCS (α = .755). Neither CRT scores (B = -.025, p = .617) nor PBCS scores (B = -.019, p = .277) predicted participants' responses to the subtraction question. A second analysis introducing familiarity with the CRT as a moderator showed no significant results from the interaction of CRT scores and familiarity on participants' responses to the subtraction question (B = -.048, p = .203).

[9] Non-parametric tests yielded similar results. One-sample Kolmogorov-Smirnov Test showed that post-subtraction emotion ratings did not follow a normal distribution, $D(189) = .230$, $p < .001$.

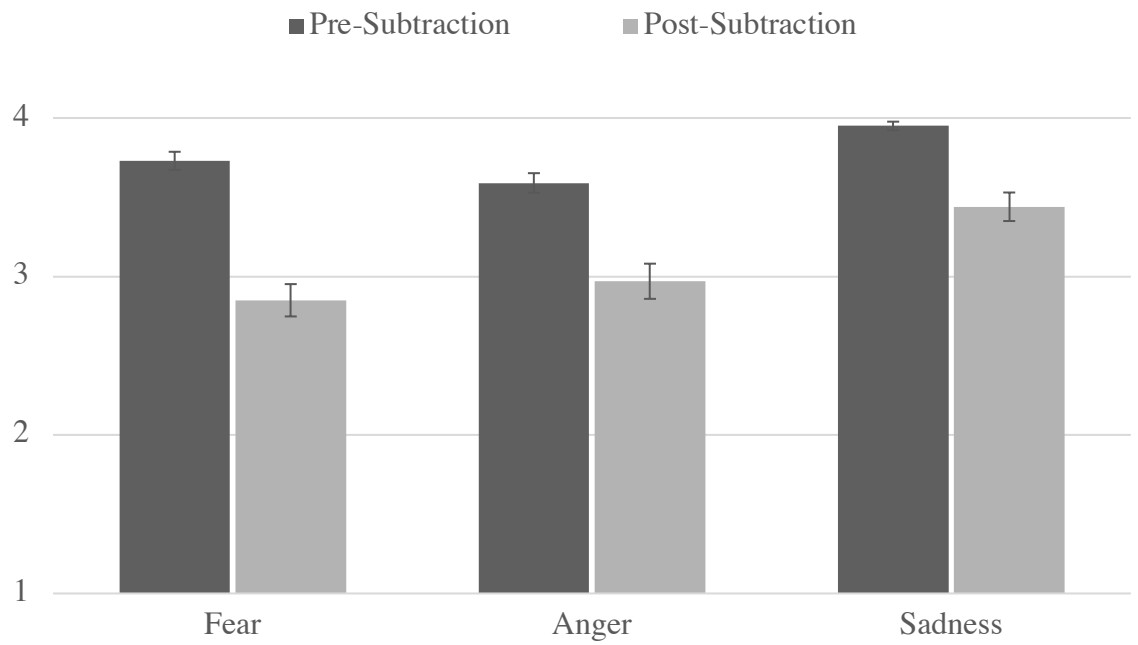

Figure 1. Mean pre- and post-subtraction emotion ratings by emotion category (Study 1). Error bars represent standard error of the mean.

4.3. Text analyses

To further strengthen confidence in the results, I conducted text analyses of participants' responses to the justification question ('Why would you still (no longer) consider yourself to be afraid (angry, sad)?'). Participants' justifications were classified by the main author in 8 categories using word-frequency queries. Afterwards, two independent coders were given written descriptions of each category and themselves coded participants' responses. Cohen's k was run to measure the agreement between their coding and the author's coding (McHugh, 2012). There was substantial agreement for most categories, with agreement percentages running from 81.98% (k = .54) to 99.10% (k = .97).

Some patterns are especially relevant. First, most participants provided relevant justifications for their responses. For example, among those participants who would *not* consider themselves afraid (angry, sad) after the subtraction, 72.09% explicitly mention that physical reactions are necessary for emotion. Thus, it seems like participants had no problems performing the task they were asked to perform. Second, none of the participants mentioned that the bodily changes remain after subtracting the *feeling* of bodily changes. Thus, unfelt bodily changes cannot explain participants' responses. Third, among participants who *would* consider themselves to be afraid (angry, sad) after the subtraction, only 8.22% mentioned the possibility of alternative bodily manifestations, and only 6.85% mentioned the possibility of internal manifestations. Thus, seems like non-subtracted bodily feelings cannot account for the results either.[10] Table 2 shows the full distribution of responses across categories.

	Emotion Post-Subtraction	
Nodes	**Yes**	**No**
Evaluation of the situation e.g. "Would be a situation that could end badly."	34.25% (50)	2.33% (1)
Emotion is mental e.g. "[…] absence any physical manifestation or fear, you're still mentally and psychologically afraid."	28.08% (41)	2.33% (1)

[10] Follow-up analyses support this claim. After excluding these two groups of participants, mean ratings for the subtraction question (*Mean* = 3.03, *SD* = .845) were still significantly higher than the scale's midpoint (2.5), $t(172) = 8.2$, $p < .001$, $d = .627$. Non-parametric tests yielded similar results. One-sample Kolmogorov-Smirnov Test showed that post-subtraction emotion ratings did not follow a normal distribution, $D(173) = .238$, $p < .001$.

Physical reaction is not necessary e.g. "You don't always show physical signs of anger, sometimes slight anger does not arouse a physical response"	24.66% (36)	0%
Physical reaction is necessary e.g. "Those are the kind of feelings that you have when you are sad or heartbroken"	0.68% (1)	72.09% (31)
Situation e.g. "The risk is still there, and it will creep on you while you least expect it"	20.55% (30)	4.65% (2)
Alternative manifestations e.g. "Those are some people's physical reactions to anger, but not everyone's."	8.22% (12)	0%
Internal vs External e.g. "On the inside I would be scared and that's what matters, not what I feel on the outside."	6.85% (10)	0%
Other e.g. "Yes"	26.03% (38)	25.58% (11)
Total	**149.32%** **(218/146)**	**106.98%** **(46/42)**

Table 2. Percentage (absolute number) of justification responses falling on each category by participants' answer to the post-subtraction question. Note that the percentages are calculated over the total number of responses, but each response might be coded at several nodes.

4.4. Discussion

In Study 1, most participants (77.3%) considered that they would *still* be afraid (angry, sad) after subtracting the bodily feelings from an emotional experience. This suggests that James' intuitions, and those of the authors that have accepted his conclusions, might be biased by their theoretical commitments (see Stich and Machery 2018). But note that, although subtracting the feeling of bodily changes didn't preclude emotion, it did have an impact on emotion ratings. Post-subtraction emotion ratings were significantly lower than pre-subtraction emotion ratings. This suggests that, in line with Disembodied accounts of emotion, bodily changes merely modulate (vs. determine) the experience of emotion. However, Embodied theories might also be able to accommodate these results.

Proponents of Embodied theories might posit that (1) participants failed to subtract the feelings of bodily changes from the experience of emotion, or that (2) participants are not sensitive enough to their own bodily states. As it has already been mentioned (§4), the Subtraction Argument demands both reflection and interoceptive awareness. The results, however, do not seem to support these explanations. Against (1), there was no relationship between participants' responses to the subtraction question and their tendency to engage in cognitive reflection. Furthermore, participants' justification for their responses suggests that they had no problems performing the task, and (most of them) subtracted *all* bodily changes from the experience of emotion. Against (2), there was no relationship between participants' responses to the subtraction question and their interoceptive awareness.

Altogether, the results of Study 1 suggests that bodily changes are *not* essential to emotion. However, one could argue that there is an important distinction to be made here. In particular, the distinction between hypothetical and actual emotions. Although *hypothetical* emotions (the ones studied in this study) could exist without bodily feelings, *actual* emotional experiences

might necessarily involve bodily feelings. In the next section, I will present the results of a follow-up study designed to address this worry.

5. Study 2

In order to test whether the results of Study 1 hold for *actual* (vs. hypothetical) emotional experiences, I reran Study 1 with some important modifications. First, instead of asking participants to *imagine* an emotion, Study 2 used an autobiographical recall task to *induce* emotion in participants. Autobiographical recall is a well-validated method for inducing emotions in experimental settings. In autobiographical recall tasks, participants are asked to think about a situation in which they felt a strong emotion (e.g. fear) and describe it in detail. As they recall their previous experiences, participants come to experience emotions anew. A recent review (Siedlecka and Denson 2019) shows that autobiographical recall reliably influences subjective emotional experience as well as physiological responses such as changes in heart rate and blood pressure. Indeed, autobiographical recall and mental imagery methods have been used to investigate the bodily correlates of emotion in almost 200 studies (Siegel et al. 2018). As participants describe the situations that triggered their emotions in the past, they come to experience emotion again, as well as the associated bodily changes. Apart from using autobiographical recall to induce an actual emotion, Study 2 introduces a second important change with respect to Study 1. Instead of using James' original wording for the subtraction of bodily feelings, participants in Study 2 were asked to subtract all feelings of bodily changes. This aims to address worries regarding non-subtracted bodily changes (see §4.3),[11] or a

[11] For recent studies on people's association between emotions and bodily changes, see Rimé et al. (1990) and Nummenmaa et al. (2014)

possible effect of the somewhat archaic language used in James' original formulation of the Subtraction Argument.

5.1. Method

99 participants were recruited through Amazon Mechanical Turk and completed the survey for a monetary payment. They were randomly assigned to one out of three cases (Fear, Anger, Sadness). 19 participants did not complete the survey and 7 participants did not report emotion after the induction. These participants were excluded, leaving a final sample of 73 participants (41 male, 32 female, 1 other, M_{age} = 35.32, SD = 10.55, age range 23-69). Post-hoc power analyses using G*Power showed that the study had 99% power to detect an effect of d = .5.

Participants first completed the emotion induction task. As in Study 1, I will describe the procedure for the Fear case. First, participants were told to 'Please think of a situation that made you extremely afraid. Vividly recall what happened and how you felt, and write it in the box below. Take at least two minutes to give as much detail as possible'. Afterwards, participants were asked to 'Please indicate to what extent you are currently feeling the following:' (1) 'fear' rated on a scale from 0 ('not at all') to 6 ('very much'), and (2) 'bodily changes (heart pounding / muscles tensing / blushing / other)' also rated on a single scale from 0 ('not at all') to 6 ('very much'). The same bodily changes, but a different emotion, were presented in the Anger and Sadness cases. Then, participants went through the subtraction task. Participants were told to 'Now imagine that you don't feel any changes in your body', and answered the question 'Would you still consider yourself to be afraid?' on a scale from 0 ('not at all') to 6 ('very much').

5.2. Results

After the induction, most participants reported feeling bodily changes (*Mean* = 2.73, *SD* = 1.694).[12] However, 11 out of the 80 participants in the study (13.7%) reported feeling emotion (ratings higher than 0) but no bodily changes.

The results regarding emotion ratings replicated those in Study 1. Pre-subtraction emotion ratings (*Mean* = 4.05, *SD* = 1.487) were significantly higher than post-subtraction emotion ratings (*Mean* = 3.45, *SD* = 1.743), $p = .005$, $d = .368$,[13] but post-subtraction emotion ratings (*Mean* = 3.45, *SD* = 1.743) were significantly higher than James' expected response (0 – 'not at all'), $p < .001$, $d = 1.979$ (see Figure 2).[14] Most participants (93.5%) gave ratings higher than 0 in the post-subtraction question.

[12] Mean bodily feeling ratings were not significantly different for fear (*Mean* = 3.00, *SD* = 1.651), anger (*Mean* = 2.46, *SD* = 1.567), and sadness (*Mean* = 2.75, *SD* = 1.567), $F(2, 70) = .614$, $p = .544$.

[13] Non-parametric tests yielded similar results. Wilcox signed-rank test showed that pre-subtraction emotion ratings (*Median* = 4) were significantly higher than post-subtraction emotion ratings (*Median* = 3), $Z = -2.699$, $p = .007$.

[14] Non-parametric tests yielded similar results. One-sample Kolmogorov-Smirnov Test showed that post-subtraction emotion ratings did not follow a normal distribution, $D(62) = .118$, $p = .031$.

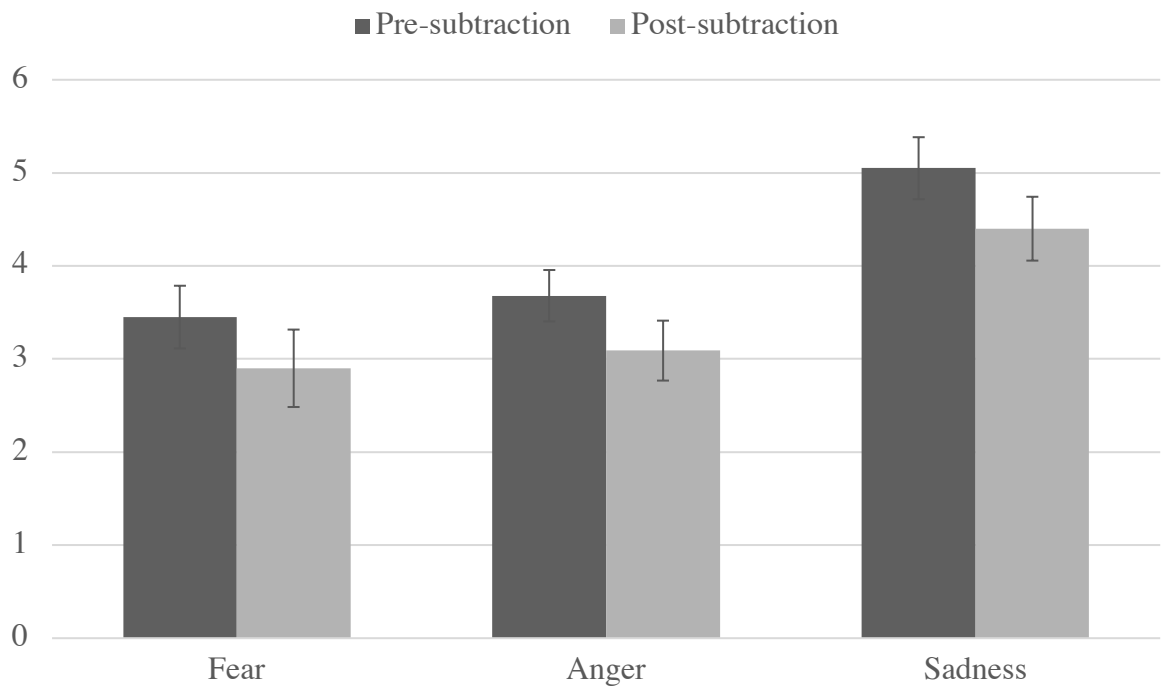

Figure 2. Mean pre- and post-subtraction emotion ratings by emotion category (Study 2). Error bars represent standard error of the mean.

5.3. Discussion

The results of Study 2 suggest that, even for actual emotions, subtracting the feeling of bodily changes does not lead to the absence of emotion. Most participants (93.5%) answered that they would still consider themselves to be afraid (angry, sad) after the subtraction. Furthermore, 11 out of 80 participants in the study (13.7%) reported experiencing emotion but not bodily changes *before* the subtraction (right after the Autobiographical Recall task). This suggests that, contrary to objections, actual emotional experiences do not necessarily involve bodily feelings.

6. General Discussion and Conclusions

As we have seen, the Subtraction Argument is a central line of support for the claim that bodily changes are essential to emotion. Authors hold opposite intuitions regarding this thought experiment, sustaining a fundamental divide in emotion theory: the one between Embodied and Disembodied theories. In order to advance this debate, I conducted a series of studies testing people's intuitions regarding the Subtraction Argument. In Study 1, participants were told to imagine an emotional experience and afterwards subtract the bodily feelings from it. Against Embodied theories, most participants reported that their emotions would persist in the absence of bodily feelings. Further analyses showed that the results cannot be accounted for in terms of participants' reflection (as measured by the CRT), introspective capacities (as measured by the PBCS), or failure to subtract bodily feelings from the experience of emotion (as shown by text analyses of participants' justifications of their responses). In Study 2, the results were replicated with regards to *actual* instead of *hypothetical* emotions. Instead of making participants imagine an emotional experience (as in Study 1), this study used an emotion induction task to elicit an actual emotional experience in participants. Again, most participants reported that their emotions would persist in the absence of bodily feelings. Furthermore, some participants indicated that they were experiencing emotion but not bodily feelings even *before* the subtraction. Overall, the results of both studies suggest that people's emotional experiences do not depend on bodily changes. Contrary to Jamesians' claims, the Subtraction Argument *denies* the claim that bodily changes are essential to emotion.

The results presented here fit well with extant evidence regarding emotional experience in cases of impaired interoception due to, e.g. spinal cord injury. Because patients with spinal cord injuries lack bodily feedback, these cases constitute a non-hypothetical equivalent to the Subtraction Argument. In the Subtraction Argument, we are told to *hypothetically* eliminate the

feelings of bodily changes from an emotional experience. After suffering a spinal cord injury, feelings of bodily changes are *actually* eliminated (to different degrees, depending on the position and severity of the injury). Some studies have found that patients with spinal cord injuries report decreases in their experience of emotion after the injury, with larger decreases for more severe or higher injuries (Hohmann 1966). However, other studies have found no decreases in emotional experience even after complete cervical injuries (Cobos et al. 2002). The study of other cases of impaired interoception, such as pure autonomic failure, has yielded similar results (for a comprehensive review, see Laird & Lacasse, 2014).

In order to defend their accounts from the evidence against it, proponents of Embodied theories of emotion have put forward different auxiliary hypotheses. Most importantly, it has been claimed that the brain can *simulate* the feeling of bodily changes in the absence of real bodily feedback (Damasio 1994; MacCormack and Lindquist 2017). Thus, spinal cord injuries do not preclude (simulated) feelings of bodily changes. This move seems to assume that, if people experience emotion in the absence of bodily feedback, they must feel bodily changes in some other way. In other words, that people report emotion only if they feel bodily changes. The results presented in this paper suggest that this is not the case, and thus challenge extant explanations for the persistence of emotion in cases of impaired interoception. Instead of assuming that spinal cord injury patients still experience simulated bodily feelings, we should perhaps concede that emotional feelings persist without bodily feelings.

Finally, it is important to note that the results here concern people's intuitions. One might be reluctant to take people's intuitions about emotion as evidence for or against emotion theories. After all, laypeople might be massively wrong about the nature of emotion. In the introduction, we have seen how authors appeal to our intuitions to support their claims about the nature of emotion. But maybe they should not. Emotion theories deal with what emotions really are, not

with people's understanding of emotions. Thus, emotion theorists might be better off by just ignoring people's intuitions. However, this move comes at a high cost. First, self-report is the only way to access the content of people's experience of emotion (Quigley et al. 2014). Second, as it has already been noted (Scarantino 2012), the object of study or *explanandum* of emotion theories is people's emotional experiences. A theory of emotion should account for the things that people refer to when they say we are angry, sad or afraid. Thus, if we ignore people's reports, we might establish a definition of emotion that points to very different phenomena. In other words, we might end up with a theory that is not a theory of emotion, but rather a theory of something else.

The results presented in this paper show that people report anger, sadness, and fear in the absence of bodily feelings. If Embodied theories want to account for people's emotional experiences, they have to *explain away* these results. Here, I have considered some possible ways of doing so, from participants' psychological traits to their failure to perform the subtraction task (see above). The data, however, does not support any of these explanations. Future studies should further investigate ways in which Embodied theories can account for this data. As things stand now, the results here add up to previous evidence against the claim that bodily changes are essential to emotion (see Reisenzein and Stephan 2014), and challenge the explanations used by proponents of Embodied theories to account for some of those previous findings (in particular, the ones regarding emotion in cases of impaired interoception). Overall, the evidence suggests that bodily changes are *not* essential to emotion. This poses a problem for Embodied theories of emotion, and in turn favors Disembodied theories, which seem better suited to account for people's emotional experiences.

References

Arnold, MB. 1960. *Emotion and personality*. New York: Columbia University Press.

Barlassina, Luca, and Albert Newen. 2014. The Role of Bodily Perception in Emotion: In Defense of an Impure Somatic Theory. *Philosophy and Phenomenological Research* 89: 637–678. doi:10.1111/phpr.12041.

Brun, Georg. 2017. Thought Experiments in Ethics. In *The Routledge Companion to Thought Experiments*, ed. Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown, 195–210. New York: Routledge. doi:10.4324/9781315175027-12.

Cobos, Pilar, Marı´a Sánchez, Carmen Garcı´a, Marı´a Nieves Vera, and Jaime Vila. 2002. Revisiting the James versus Cannon debate on emotion: startle and autonomic modulation in patients with spinal cord injuries. *Biological Psychology* 61. Elsevier: 251–269. doi:10.1016/S0301-0511(02)00061-3.

Cochrane, Tom. 2017. The Double Intentionality of Emotional Experience. *European Journal of Philosophy*. doi:10.1111/ejop.12233.

Damasio, Antonio. 1994. *Descartes' Error. Emotions, Reasons and the Human Brain*. New York: Avon Books.

Deonna, Julien A., and Fabrice Teroni. 2017. Getting Bodily Feelings Into Emotional Experience in the Right Way. *Emotion Review* 9: 55–63. doi:10.1177/1754073916639666.

Deonna, Julien A, and Fabrice Teroni. 2012. *The emotions : a philosophical introduction*. New York: Routledge. doi:10.1016/B978-0-12-415956-3.00001-4.

Foot, P. 1967. The problem of abortion and the doctrine of double effect.

Frederick, Shane. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19: 25–42. doi:10.1257/089533005775196732.

Goldie, Peter. 2000. *The Emotions: A Philosophical Exploration*. New York: Oxford University Press. doi:10.2307/3182582.

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

Helm, Bennett W. 2009. Emotions as Evaluative Feelings. *Emotion Review* 1: 248–255. doi:10.1177/1754073909103593.

Hohmann, George W. 1966. Some Effects of Spinal Cord Lesions on Experienced Emotional Feelings. *Psychophysiology* 3: 143–156. doi:10.1111/j.1469-8986.1966.tb02690.x.

Hufendiek, Rebekka. 2016. *Embodied emotions : a naturalist approach to a normative phenomenon*. New York: Routledge.

James, William. 1884. What is an Emotion ? *Mind* 9: 188–205. doi:10.1093/mind/LI.202.200.

James, William. 1890. *Principles of psychology Vol. 2*. New York: Henry Holt & Co.

Kreibig, Sylvia D. 2010. Autonomic nervous system activity in emotion: A review. *Biological Psychology* 84. Elsevier B.V.: 394–421. doi:10.1016/j.biopsycho.2010.03.010.

Laird, James D. 2007. *Feelings: The perception of self. Feelings: The Perception of Self*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195098891.001.0001.

Laird, James D., and Katherine Lacasse. 2014. Bodily Influences on Emotional Feelings: Accumulating Evidence and Extensions of William James's Theory of Emotion. *Emotion Review* 6: 27–34. doi:10.1177/1754073913494899.

Lange, C. G. 1922. The emotions: A psychophysiological study. In *The Emotions*, ed. Knight Dunlap, 33–90. Baltimore: Williams & Wilkins Company.

Lazarus, RS. 1984. On the primacy of cognition. *American Psychologist* 39: 124–129.

MacCormack, Jennifer K., and Kristen A. Lindquist. 2017. Bodily Contributions to Emotion: Schachter's Legacy for a Psychological Constructionist View on Emotion. *Emotion Review* 9: 36–45. doi:10.1177/1754073916639664.

Miller, Lynn C., Richard Murphy, and Arnold H. Buss. 1981. Consciousness of body: Private and public. *Journal of Personality and Social Psychology* 41: 397–406. doi:10.1037/0022-3514.41.2.397.

Mitchell, Jonathan. 2018. The Irreducibility of Emotional Phenomenology. *Erkenntnis*. Springer Netherlands. doi:10.1007/s10670-018-0075-8.

Nummenmaa, Lauri, Enrico Glerean, Riitta Hari, and Jari K Hietanen. 2014. Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the United States of America* 111: 646–51. doi:10.1073/pnas.1321664111.

Nussbaum, Martha. 2001. *Upheavals of thought*. Cambridge: Cambridge University Press. doi:10.5840/gfpj200223213.

Prinz, Jesse. 2004. *Gut reactions: A perceptual theory of emotion*. New York: Oxford University Press.

Quigley, Karen S, Kristen a Lindquist, and Lisa Feldman Barrett. 2014. Inducing and Measuring Emotion and Affect: Tips , Tricks , and Secrets. In *Handbook of Research Methods in Social and Personality Psychology*, ed. Harry Reis and Charles Judd, 220–252. New York: Cambridge University Press. doi:10.1017/CBO9780511996481.

Reisenzein, Rainer, and Achim Stephan. 2014. More on James and the physical basis of emotion. *Emotion Review* 6: 35–46. doi:10.1177/1754073913501395.

Rimé, Bernard, Pierre Philippot, and Daniela Cisamolo. 1990. Social Schemata of Peripheral Changes in Emotion. *Journal of Personality and Social Psychology* 59: 38–49.

doi:10.1037/0022-3514.59.1.38.

Rossi, Mauro, and Christine Tappolet. 2018. What kind of evaluative states are emotions ? The attitudinal theory vs . the perceptual theory of emotions. *Canadian Journal of Philosophy* 00. Routledge: 1–20. doi:10.1080/00455091.2018.1472516.

Scarantino, Andrea. 2012. How to Define Emotions Scientifically. *Emotion Review* 4: 358–368. doi:10.1177/1754073912445810.

Scarantino, Andrea, and Ronald de Sousa. 2018. Emotion. In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. doi:10.1111/1467-9973.00225.

Schachter, Stanley. 1964. The Interaction of Cognitive and Physiological Determinants of Emotional State. *Advances in Experimental Social Psychology* 1: 49–80. doi:10.1016/S0065-2601(08)60048-9.

Siedlecka, Ewa, and Thomas F. Denson. 2019. Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review* 11: 87–97. doi:10.1177/1754073917749016.

Siegel, Erika H., Molly K. Sands, Wim Van den Noortgate, Paul Condon, Yale Chang, Jennifer Dy, Karen S. Quigley, and Lisa Feldman Barrett. 2018. Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin* 144: 343–393. doi:10.1037/bul0000128.

Solomon, Robert C. 2003. *Not passion's slave: Emotions and choice*. New York: Oxford University Press. doi:10.1177/1754073909345543.

Stich, Stephen, and Edouard Machery. 2018. A Possible Future For Philosophy. *The Philosophers' Magazine*: 98–100. doi:10.5840/tpm20188026.

Tappolet, Christine. 2016. *Emotions, Values, and Agency*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780199696512.001.0001.

Waldmann, Michael R., Jonas Nagel, and Alex Wiegmann. 2012. Moral Judgment. *The Oxford Handbook of Thinking and Reasoning*: 274–299.

Zajonc, RB. 1984. On the primacy of affect. *American Psychologist* 39: 117–123.