Kris Goffin

BETTER SCARED THAN SORRY:

The Pragmatic Account of Emotional Representation

ABSTRACT

Some emotional representations seem to be unreliable. For instance, we are often afraid when there is no danger present. If emotions such as fear are so unreliable, what function do they have in our representational system? This is a problem for representationalist theories of emotion. I will argue that seemingly unreliable emotional representations are reliable after all. While many mental states strike an optimal balance between minimizing inaccurate representations and maximizing accurate representations, some emotional representations only aim at maximizing accuracy. They detect important phenomena such as danger based on little evidence, which will lead to a lot of false alarms. When it matters, however, these emotional representations will detect danger and other important phenomena. Often, one is better scared (and wrong) than sorry.

KEYWORDS: Philosophy of Cognitive Science; Emotion; Philosophy of Perception; Information-Based Accounts of Mental Content; Fear

Kris Goffin

BETTER SCARED THAN SORRY:

THE PRAGMATIC ACCOUNT OF EMOTIONAL REPRESENTATION

I. THE EMOTIONAL UNRELIABILITY PROBLEM

When walking through the jungle, Jack suddenly hears an unexpected sound. He is startled and afraid. However, it turns out that all he heard was a branch moving in the wind. Therefore, he had no reason to be afraid. This sort of emotional misfiring seems to be much more common than perceptual illusions.

Emotions such as fear seem to provide us with information about the world. Fear seems to tell us that there is danger present. Jack learns from fear that the sound in the jungle is something dangerous. Another way of describing this would be to say that fear represents danger. I will call this representational aspect of emotion "emotional representation", without claiming that an emotion can be reduced to its representational capacity.[1] Also, I do not imply that the phenomenology of an emotional experience is exhausted by its representational capacity.

---

[1] It is possible that our folk psychological concept "emotion" refers to a set of mental states, rather than one monolithic representation. In fact, some emotion scientists claim that an emotion consists of multiple components See e.g. (LeDoux, 2014) or (Moors, 2017). Depending on which emotion theory one defends, emotions might include different components, such as directive content (or "action tendencies"), bodily feelings, facial expressions, behaviour and higher cognitive thoughts. If you accept such a view of emotion, you should note that "emotional representation" refers to the aspect of emotion that is representational and is distinct from higher cognition. When I talk about "emotion" in this paper, I mean "emotional representation", while remaining neutral on the definition and ontology of "emotion" as a folk psychological concept.

Emotional representations, however, do not seem to be a very reliable source of information. In his seminal work on reliability, Goldman (1986, p. 43) uses beliefs caused by feelings and by being in a bad mood as paradigm examples of unreliability. A reason why emotional representation might be treated as unreliable is that it often misfires. For instance, we could consider the fear of the sound Jack hears in the jungle a false alarm.

For another example, suppose that Lucy is watching a 3D movie when suddenly a 3D snake jumps towards her. She is startled, although she is fully aware that there is no real snake there. We often respond emotionally to things that do not exist, such as fictional movie characters.

In addition, emotional representations could sometimes be considered to be exaggerations of the facts. For instance, when we are angry at someone who has offended us, we often tend to exaggerate how big the offence was. When we have calmed down, we realize that what we were angry about does not matter that much. Disgust is another example of an emotion which does not seem to be very reliable. When we are disgusted, we tend to exaggerate the danger of contagion (see Rozin et al., 1986). For instance, being disgusted by a bit of mould in one's kitchen might cause one to refuse to eat any food that has been in that kitchen.

The seeming unreliability of emotional representation creates a problem for philosophers of mind. Emotional representations are unarguably an important aspect of the human mind. Nevertheless, if they are so unreliable, what do we need them for? As emotional representations do not seem to be very good at representing reality, why could we not just rely on perception and reasoning alone? What function, then, do emotional representations have in our representational system? Let us call this the "emotional unreliability problem".

This problem does not refer to the "(ir)rationality" of emotion: it can be argued that it is "rational" to be afraid in the jungle when you hear an unexpected sound because it "makes sense" to do so. This paper is, however, not about the rationality of emotion.

The "emotional unreliability problem" is a problem for representationalism about emotion. Suppose that emotional representations are not reliable. If this is true, then it becomes hard to identify the function of emotional representation in our representational system.

A common way to think of mental representations is as those mental states that have the function to reliably correlate with what they represent.[2] Reliable correlation does not imply infallibility, because what can represent can also misrepresent. But if a mental state does not reliably correlate with what it is supposed to represent, it seems that this mental state is not what we traditionally would call a mental representation. Suppose that all maps of London do not reliably correlate with the geographic facts about London. You would wonder why people would use these maps, and you would even question whether maps of London are actual maps. Because reliability is closely tied to the notion of representation, a mental state which "unreliably correlates" with what it is supposed to represent almost seems paradoxical.

Applied to emotions, one can say that the function of fear is to reliably correlate with "danger". When danger occurs, one is supposed to be afraid. Tappolet (2016, pp. 41–42) states that emotions misfire more often than not. Many instances of fear are false alarms. Therefore, we

---

[2] See, for instance, the work of Fred Dretske (1981, 1986, 1988).

have reason to give up thinking of emotion in terms of reliable correlation altogether. For this reason, Tappolet rejects the reliable covariation theory of emotion. This criticism is aimed at Prinz (2004) who defends a Dretskean account of mental representation, and thinks of representation in terms of reliable covariation (see also Price, 2015 and Scarantino, 2014). Tappolet's criticism is related to a more general criticism of a Dretskean account of representation, namely that it cannot account for states which produce a lot of false alarms (see also Burge, 2010, pp. 292-308). This criticism should be taken seriously by those who understand mental representation in terms of reliable covariation.

If there are no stable and reliable links with the outside world then, according to Dretske (1981), meaning (or mental representation) and knowledge becomes impossible (see also Godfrey-Smith, 1998, pp. 116-121). I will not focus on knowledge and related epistemological questions in this paper. I will rather focus on mental representation.

In this paper, I will present a solution to the emotional unreliability problem. First, I will clarify a couple of things about the emotional unreliability problem. Then, I will distinguish between two kinds of reliability, after which I will present and defend the pragmatic account of emotional representation as a solution to the emotional unreliability problem. In the final section, I will consider some objections.

II. SOME CLARIFICATIONS

This section is mostly concerned with clearing the ground. There are two things that need clarification before we proceed. First, I do not think that the emotional unreliability problem

applies to all phenomena we call "emotions". Secondly, I will say more about how emotions provide us with information about the world.

The emotional unreliability problem does not apply to all emotions. Most examples I gave above, as well as the examples used by Tappolet (2016, pp. 41–42), are about a limited set of emotions. Fear and disgust seem to misfire a lot, but it is hard to see why more complex emotions such as melancholy or positive emotions such as joy would be unreliable in the same sense.

When providing an answer to the emotional unreliability problem, I will explain why some emotions seem unreliable, while others do not. Until I address this point I will use "emotional representation" to refer to those emotional representations which seem unreliable because they present us with a lot of false alarms, such as the emotional representations involved in the examples I gave above.

I take it to be uncontroversial enough to suppose that emotional representations provide us with information about the world in some sense, but I would like to say a bit more about the kind of information that is conveyed. First of all, emotional representations can be more or less intensely positive or negative. One way of describing this phenomenon is by saying that emotional representations represent their objects as more or less good or bad.[3] If one fears a lion, for instance, one experiences it as "bad" in some sense. More specifically, one experiences the lion as "dangerous", by means of emotional representation. Emotional representations

---

[3] I am trying to explain what emotions represent in an intuitive way. I do not mean that emotions literally represent the thin values of "good" and "bad".

also vary in intensity. Hearing your favorite song on the radio might give you joy, but seeing your newborn child will almost certainly give you a more intense experience of joy because it signals something of greater importance. Emotions seem to be relational as well. One represents objects as, for example, "bad for me", and not necessarily "bad for everyone".[4] Another way of describing this is to say that emotional representations represent their objects as relevant to us in a good or bad way.[5] In short: emotional representations provide us with information about whether, and to what extent an object, is "good" or "bad" for us.[6]

As it stands, this characterization of what is represented in emotional representation is too broad. Arguably, our emotions can represent only a limited amount of different types of "good" and "bad", such as being dangerous or being lovable. A less colloquial way of referring to types of good and bad is as evaluative properties. The evaluative properties associated with emotions are often called "formal objects" (Kenny, 1963). A formal object can be understood as the evaluative property that makes an emotion intelligible, that individuates different emotional kinds and also provides the correctness conditions of an emotion. These three functions are in the Dretskean framework I presuppose attributed to the content of a representation. I thus presuppose that formal objects are part of the content of emotional representations. It is, however, not my aim to give an account of the nature of formal objects. We would end up discussing the metaphysics of evaluative properties, and this is a topic for

---

[4] Before an emotional representation occurs, the object of that representation arguably needs to be represented by some prior mental state, such as a perception, belief, or imagination. Deonna and Teroni (2012) call this an emotion's cognitive base. The represented object might be something in the past, as in regret. It might be something in the future, as in anxiety.

[5] See e.g. (De Sousa, 1987; Goldie, 2007; Greenspan, 2006; or Griffiths, 1997).

[6] See (Scarantino, 2010) for a clear description of what it means to say that emotions represent.

another paper.[7] For this paper, I do need to stipulate that I presuppose that emotional

representations represent a limited range of evaluative properties, which most philosophers of

emotions call the formal objects of emotion.

I also want to remain neutral about which kinds of mental state an emotion is. There is an

ongoing debate on whether emotional representations are judgements or belief-like states

(Solomon, 2003; Nussbaum, 2001), perceptual or perception-like states (Döring, 2007), or

another kind of mental state, such as a specific attitude (Deonna and Teroni, 2012). Thinking

of representations in terms of reliable correlation has been often associated with theories of

perceptual content. However, the idea that emotional representations reliably correlate with

what they represent does not presuppose nor entail that emotion is perceptual.

In this paper, I presuppose the Dretskean framework that understands mental representation

as reliable covariation (Dretske, 1981, 1986, 1988; see also Price, 2015 and Prinz, 2004)). The

Dretskean theory of mental representation I appeal to is meant to be a theory of all mental

content, regardless of which kind of mental state it is the content of. It should not matter

whether it is perception, belief or emotion. Therefore, this paper is not concerned with

describing what distinguishes emotions from other mental states. It is rather part of a

unificatory project that aims to capture the nature of all representations, within the Dretskean

---

[7] Psychologists have described these evaluative properties in relation to one's "goals" and "needs", which one might call matters of "personal significance" (For an overview, see e.g. Smith & Kirby, 2009). In the philosophy of emotion, there is an ongoing debate regarding what personal significance is. Price (2015, pp. 116–131) distinguishes between the interest-based account and the preference-based account. The interest-based account conceptualizes significance in terms of a list of objective "goods" or "interests", such as "health; security; adequate material resources; good social status; autonomy; good social relationships; intellectual stimulation" (Price, 2015, pp. 117–118). The preference-based account holds that personal significance should be thought of in terms of preferences, such as desires, values, likes, and dislikes (see also Döring, 2007; Solomon, 2003). Neo-sentimentalist accounts describe the ontology of values in terms of emotional fittingness (see for instance D'Arms & Jacobson, 2000).

framework. The emotional reliability problem is of great concern for such a unificatory project. As a response to this problem I will argue that some emotional representations have a different "style" of being reliable. These emotional representations, however, are like all other representations to be defined in terms of reliable covariation.

This might cause one to think that this paper contradicts recent attempts to distinguish emotional representation from other kinds of mental representation (Mitchell, 2019; Müller, 2019). I, however, believe that the claims I make in this paper are to some extent compatible with these theories. They might complement each other. Mitchell and Müller focus on various aspects of emotion, such as emotional phenomenology (i.e. what it feels like to have an emotional experience from the first-person perspective of the experiencer) and the vehicle of the emotional representation (i.e. the aspect of the mind that "has" a particular content). These are aspects of emotion I remain neutral on in this paper. It is plausible that one aspect which makes emotions unique is its phenomenology. It might even be the case that the particular relation between content and phenomenology is what makes emotions different from other mental states.

The account of mental representation by the "early Dretske" (1981, 1986, 1988), which this paper presupposes, does not mention phenomenology at all. This paper follows this way of thinking: I will discuss mental representation, while being neutral about how this relates to phenomenology. "Later Dretske" (1995) did make some claims about how mental representation and phenomenology relate. Dretske argued that phenomenology supervenes on content. This view is related to the idea of phenomenal transparency, which Mitchell (2019) argues, may be applicable to perception but not to emotion. I do not have the space to discuss

this in detail in this paper, but the basic idea is that the phenomenology of, for instance, seeing red is nothing over and above having a representation of something as red. Phenomenology can thus be reduced to content: if you have a theory of content, you get the phenomenology for free. This might work for the phenomenology of perceiving properties such as "red", but the phenomenology of fear arguably cannot not be reduced to a representation of something as dangerous, because the phenomenology of fear arguably also consists of bodily sensations. I distance myself from "later Dretske" and do not defend the claim that phenomenology is reducible to content or that emotional experience is "transparent".

Because I am neutral on phenomenology, the claims of this paper are consistent with most theories of emotional phenomenology. The idea of emotional content as reliable covariation is, however, in contradiction with theories that describe emotional phenomenology as "presentational" or "disclosing" value, as the aforementioned Tappolet (2016) does. Tappolet therefore has to be critical of Dretskean theories. On this account, emotions are a direct perception of value; an openness to the world: a window through which the world is revealed. This is in contradiction with the idea of emotional representations of the world which are defined as reliable covariations of what they represent. As this implies that emotional representation can misfire and misrepresent, this is in contradiction with the claim that emotions disclose value. Müller (2019) is an ally of mine to the extent that he attacks this idea of perceptual disclosure. He works from a phenomenological perspective. I work from a different angle but our perspectives are, I believe, complementary.

I think that emotions might differ from other mental states in various different ways: they may be phenomenologically different, or in emotional states phenomenology and content may be

related in a particularly distinctive way, or it may be the case that a particular vehicle or mental attitude distinguishes emotion from other mental states. In this paper, however, I just want to defend the idea that content is to be defined as reliable covariation and that emotional representation is not a counterexample to this claim. I believe that one can isolate a theory of what makes a mental content a mental content from claims about how mental states differ as well as from claims about phenomenology. This is all in the spirit of Dretske's earlier work (1981, 1986, 1988).

One might of course disagree with the idea that a defence of emotional representation as reliable covariation can be isolated from all these other aspects of emotion. I will, however, presuppose this Dretskean framework and consider what I take to be a serious objection to this way of thinking about mental content: the emotional unreliability problem. The relation between content and phenomenology might be the focus of a different objection to this Dretskean framework but this is beyond the scope of this paper.

In this section, I wanted to clarify what I mean by the statement that emotions seem to provide us with some information about the environment. Having done this, we can now reconsider the emotional unreliability problem: emotional representations seem to be unreliable, as they often present us with false alarms. What function, then, do they have in our representational system? I will argue that the pragmatic account of emotional representation shows that seemingly unreliable emotions are reliable in a distinctive way. First, I will discuss the philosophical framework that is able to distinguish between two kinds of reliability.

III. TWO KINDS OF RELIABILITY

There is an intuitive sense that emotional representation is more closely related to the primitive urge to survive than to truth. This might be a possible solution to the emotional reliability problem: emotional representations keep us alive, although they are not reliable representations of reality. However, the distinction between "truth" as a goal and "survival" as a goal can be deflated. For instance, Quine (1969) famously says that organisms that do not represent reality accurately would probably die before they can reproduce. It is a mainstream view that representing correctly is advantageous in the light of survival. Many philosophers agree with Quine and consider this as an evolutionary argument that guarantees that our cognitive capacities are aimed at representing reality accurately (Dennett, 1987; Papineau, 1987). If one presupposes that our representational systems are aimed at accurately representing reality, emotions seem to be terrible at doing what they are supposed to do.[8]

Stich (1990), on the other hand, argues that sometimes it is selectively advantageous to misrepresent. Often one is better "safe" (but false) than "sorry":

> A very cautious, risk-aversive inferential strategy – one that leaps to the
> conclusion that danger is present on very slight evidence – will typically
> lead to false beliefs more often, and true ones less often, than a less
> hair-trigger one that waits for more evidence before rendering a
> judgment. Nonetheless, the unreliable, error-prone, risk-aversive

---

[8] One might reject this Quinean way of thinking on other grounds. Different theories of emotion might provide different solutions to this problem. However, as I want to defend the idea that content is reliable covariation, I need to take this line of thinking very seriously.

strategy may well be favored by natural selection. For natural selection

does not care about truth; it cares only about reproductive success. And

from the point of view of reproductive success, it is often better to be

safe (and wrong) than sorry. (Stich, 1990, p. 62)

An organism that is sensitive to danger might misrepresent things as dangerous that really are

not. However, when it matters, the organism will represent the possible killer as dangerous. It

will act upon this representation and survive. In contrast, an organism that is not sensitive to

danger might never misrepresent something as dangerous, but perhaps it will not detect some

present and real danger. Misrepresenting something as dangerous is not as bad as not

representing the danger that could kill you.

One might think that Stich (1990) implies that it is a virtue that our representational system is

not fully reliable. This is, however, not true if we distinguish between two kinds of reliability,

as Godfrey-Smith (1991, 1998) does. A process that produces beliefs can be reliable if it

maximizes true beliefs and it can be reliable if it minimizes false beliefs. Following Godfrey-

Smith (1998), we can call these conceptions of reliability Jamesian reliability (maximizing true

beliefs) and Cartesian reliability (minimizing false beliefs). Godfrey-Smith (1991, 1998) argues

that, because our senses are imperfect and the environment is not epistemically co-operative,

it is impossible to believe truly all the time. Still, a large number of mental representations

reliably correlate with the world. An explanation for this is that in a representational system,

there is always a trade-off between the two kinds of reliability. In every context, there is a

specific optimal balance between the two kinds of reliability which is evolutionarily

advantageous. In a hostile environment where there are many dangers, it is good for an

organism to maximize true beliefs about danger. An organism that is sensitive to danger is better able to survive in this environment. On the other hand, in an environment where there is an abundance of water but not all of it is drinkable, the organism might want to minimize false beliefs about what is drinkable. Otherwise, it will get sick quite soon. Of course, these two examples are oversimplifications. In reality, the trade-off between the two kinds of reliability is very complex.

The names Godfrey-Smith (1998) gives to these two kinds of reliability obviously refer to the philosophical claims of William James (most importantly James, 1899) and René Descartes (most importantly Descartes, 1724).[9] Nevertheless, the idea of distinguishing between the two actually comes from psychophysical applications of signal detection theory, in which Cartesian reliability is referred to as "conservative bias" and Jamesian reliability as "liberal bias".[10]

Traditionally, philosophers of mind and epistemologists have conceptualized the reliability of our mental states in Cartesian terms (Dretske, 1981; Fodor, 1975; Goldman, 1986). According to the Cartesian, our representational system is aimed at mirroring reality as correctly as possible. It is better to miss out on endorsing a true claim than to misrepresent the world. Stich and Godfrey-Smith both challenge this idea. Stich argues that we are often better safe (but wrong) than sorry. In fact, he argues that the mind is more Jamesian than Cartesian. Godfrey-Smith, on the other hand, convincingly shows that in a representational system, there

---

[9] In (Godfrey-Smith, 1991), he uses the slightly more confusing concepts posited by Field (1990), namely world-head reliability and head-world reliability.

[10] The basics of signal detection theory are effectively explained in (Godfrey-Smith, 1991). (For more information, see Lynn & Barrett, 2014; Macmillan & Creelman, 2004; or Swets & Green, 1964). Signal detection theory also mentions "neutral bias" in which the bias is strictly speaking neither Jamesian nor Cartesian (Lynn & Barrett, 2014).

is always a trade-off between the two kinds of reliability. Despite their differences, Stich agrees with Godfrey-Smith's idea that "it is not just producing more truths and fewer falsehoods that makes one system of reasoning better than another; the system must excel at producing certain kinds of truths and avoiding certain kinds of errors" (Stich, 1990, p. 162. n. 7).

The idea behind the trade-off between Jamesian and Cartesian reliability has been used in theories about decision-making and belief, such as Stich's and Godfrey-Smith's. However, I think we can apply it to every mental representation, including emotional representations

IV. THE PRAGMATIC ACCOUNT OF EMOTIONAL REPRESENTATION

The idea of the Jamesian-Cartesian trade-off offers a solution to the emotional reliability problem. In this section, I will argue that those emotional representations which are not reliable in the Cartesian sense, are, in fact, Jamesian-reliable. These emotional representations aim at maximizing accurate representations about relevant phenomena.[11] This is the pragmatic theory of emotional representation.[12]

It might be advantageous for an organism to have a fear response, and represent "danger" or something similar, when there is very slight evidence that something is dangerous: the sound of a branch moving in the jungle, a 3D snake jumping towards you, or a little bit of mould in

---

[11] "True belief" might sound a bit awkward in this context because it might seem to suggest that an emotional representation is, in fact, a belief. Therefore I replace "true/false belief" when talking about emotional representation with the more neutral "accurate/inaccurate representation".

[12] The "pragmatic theory of emotional representation" is a reference to Stich's (1990) "pragmatic theory of cognitive evaluation". This theory should not be confused with Tanesini's (2008) "pragmatist" theory of emotion, stressing the value and virtue of fallibilism, which is also inspired by Jamesian philosophy.

your kitchen. This will result in many false alarms, but it will keep the organism alive when it matters. In the same sense, exaggerated anger can be considered to be a maximization of accurate representations. It is better to be really angry and as a consequence overly aggressive than to underestimate the importance of the situation and be killed. In other words, all the examples I gave to suggest that emotional representations are not reliable do in fact show that these emotional representations are very reliable in the Jamesian sense.

These emotional representations are Jamesian-reliable processes that provide relevant information about the world when it matters. It matters that we represent phenomena such as danger when it occurs. Often it is more costly for an organism to miss danger than to misrepresent it. Emotional representations are often inaccurate, which is a consequence of the Jamesian strategy: to be safe rather than sorry. Often that means that one is better scared than sorry.

I have been using examples which refer to primitive emotions rooted in survival. Yet most emotions we experience in everyday life relate to phenomena which are relevant to the individual for various reasons. There are many phenomena we encounter which are relevant for our well-being, and this includes more than just the things that might get us killed.

What is relevant and how relevant it is, varies from person to person. Different people judge things to be of importance to varying degrees.[13] Past feelings of loneliness might exacerbate a person's sensitivity to cues of social exclusion.[14] The Jamesian emotional process does not

---

[13] This has been the subject of a large body of research. (See e.g. De Houwer et al., 2001, Hofmann et al., 2010).

[14] Confusingly, the ordinary language notion of "sensitivity", in the sense I am using it here, is not what sensitivity means in signal detection theory (SDT). (This footnote is a continuation of footnote 10). Sensitivity in SDT is

need much evidence to signal to its owner that somebody is laughing at her or is mad at her. A previous bad relationship might cause an individual to be hypersensitive to cues that the next boyfriend may leave her, which results in a lot of fearful false alarms. A trauma causes a person to be hypersensitive to trauma-related stimuli. When such a stimulus occurs, it is likely that a Jamesian-reliable representation will follow.

In Section II, I said that the emotional reliability problem does not apply to every emotion. And now we can see why. With most kinds of mental representations, there is always a trade-off between the Jamesian and Cartesian reliability. It is a matter of degree. In some cases, it is better to be more Jamesian, and sometimes it is better to be more Cartesian. Most emotions, like other representations, are looking for the right spot on the Jamesian-Cartesian spectrum.

However, emotions such as fear and disgust are special because their function is to represent things in the environment that you really must not miss. It is very important not to miss the danger that might kill you, or the food that might make you sick. Because of the nature of the things that these emotions represent, they are always on the Jamesian end of the spectrum. This leads them to present us with many false alarms, which may give the impression that they are unreliable. It is, however, wrong to say that they are unreliable. They are not reliable in the Cartesian sense, but they are Jamesian-reliable. The reason why this does not hold for all emotional representations is that not all emotional representations detect phenomena that in most cases are better misrepresented than missed.

---

the ability to distinguish a particular signal from particular noise. The common-sense notion of sensitivity as I am using it here has more in common with the SDT notion of bias (Lynn & Barrett, 2014).

The pragmatic account of emotional representation, however, shows how emotional representations which seem unreliable are in fact reliable in the Jamesian sense, and have a specific function in our representational system. One needs the pragmatic account of emotion if one wants to defend the reliable covariation theory of mental representation. Some emotional representations could be used as counterexamples to this theory because they do not seem to be reliable at all. Therefore it becomes unclear what their function is in a representational system. The pragmatic account of emotional representation offers a solution to the emotional unreliability problem. Consequently, we do not have to give up on the reliable covariation theory of emotional representation.

V. OBJECTIONS

I will now consider possible objections to the pragmatic account.

The first objection I want to mention is actually more a point of clarification, rather than a serious objection. One can remark that what it takes to be fully Jamesian-reliable and maximize accurate representations is to represent everything at once: one should have all the emotions all the time. This might actually be impossible, but let us for the sake of the argument presuppose that this would be possible to experience all emotions all the time. Taking the idea of Jamesian reliability to the extreme would mean that you only care about maximizing accurate representations and not all about false alarms. If you represent everything which is plausible, you will never miss anything, although you will be presented with a massive amount of false alarms. This is, however, not implied by saying that some emotions are Jamesian-reliable and not Cartesian-reliable. The claim that a more Jamesian-reliable strategy is

preferable to a more Cartesian-reliable strategy does not imply that the "most" Jamesian strategy will be the most effective. As I said, there is a spectrum going from very Jamesian to very Cartesian. I did not mean to say that the most extreme pole of that spectrum which one can imagine would be best in a situation where a more Jamesian strategy would be preferred.

There is always a trade-off between maximizing accurate representations and minimizing inaccurate representations. What determines the trade-off is the calculation of what is more costly for the organism: to misrepresent or to miss? Experiencing every emotion at once is not advantageous or useful at all. It is very costly for an organism to experience every emotion all the time, because the organism would be overwhelmed and not able to function at all. The amount of evidence according to a Jamesian strategy to represent something is significantly lower than it is when following a Cartesian strategy. But that does not mean that there is no threshold whatsoever.

Another worry one might have is that the emotional unreliability problem was already solved by Millikan's (1989, 1991) theory of mental representation.[15] Millikan argues that mental representation does not involve reliable covariation and that mental representation is instead entirely determined by biological function. One can argue that fear gives rise to a lot of false alarms because it is not meant to reliably covary with danger, it is just meant to function in such a way that it keeps us alive. Millikan's theory is thus not really a solution to the emotional unreliability problem but rather a rejection of the Dretskean framework I presuppose. It is a rejection of the idea that emotional mental representations are reliable indicators of what they

---

[15] See also (Griffiths, 1997) and (Scarantino, 2014) on Millikan's theory of representation and emotion.

represent, which is the very idea that this paper is defending. One motivation in favor of Millikan's theory is that it can account for seemingly unreliable representations such as danger representation. I hope to have shown in this paper that this seemingly unreliability does not imply that we have to give up the idea of reliable covariation entirely.

I will also provide additional reason to prefer the idea of reliable covariation over Millikan's account. Burge (2010, pp. 292-308) offers a good criticism of this theory, which I will summarize here. Millikan's account of representation states that mental content is grounded in biological function. Misrepresentation could then be explained as malfunctioning. False alarms such as the paradigm example of this paper, are also explained by referring to biological function. Because it is better in the light of evolution to be hypersensitive to danger, these false alarms are not cases of malfunctioning. Burge argues that this is a mistake. One cannot infer from the practical benefit of a single case of reliable detection failure that this serves a larger biological function. Moreover, Burge argues, functioning in successful interaction with a certain condition such as danger is not identical to accurately representing the danger. The relation between accurate representation and biological function (which is grounded in fitness) is more complex than Millikan claims, so Burge says. I agree with Burge's criticism and believe that the idea of the Jamesian - Cartesian trade-off offers a good explanation of how accurate representation and biological function relate in a more complex way. In this paper, I have argued why we should maintain the idea of representation as reliable covariation, which is, I believe, a better account of mental representation than a simple reduction to biological function. More can be said about these issues, but this is beyond the scope of the paper.

The third worry about the pragmatic theory of emotional representation is that it is not certain

that the cases I call false alarms are in fact false alarms.[16] We can take the content of fear to be "possibility of danger" rather than "occurrent danger": in fear there is always a high degree of uncertainty involved (see f.i. Gordon, 1987). If this is true, then the fear of the sound in the jungle example would not be a false alarm as there is a probability of danger. I, however, do not think that it is true that fear detects the mere possibility of danger. Suppose that Clara has the scientific knowledge and expertise to create a poison that can kill a human being within five minutes. By accident, she drinks the lethal dose of the poison she has created. Clara knows that she will die and there is nothing to do about it. It goes without saying that Clara is afraid. In this case, there is no uncertainty that the poison is dangerous: it is a fact. Clara's fear indicates occurrent danger, not the mere possibility of danger. I take Clara's fear to be an accurate representation. Common sense tells us that a branch moving in the wind, which will not cause any harm, is not dangerous. Representing it as dangerous is an inaccurate representation or a false alarm.

However, one might think that "certain danger" is in fact an "extremely high possibility of danger", if you want to push the idea of fear representing possibilities. If you follow this line of reasoning, you can say that in the case of a "false alarm" there is a much lower degree of possibility than fear has represented there to be. This is consistent with my story. Fear often represents an extremely high possibility of danger while in fact there is a very low possibility of danger. It is thus still justified to speak of "false alarms" in these cases.

Another reason why one might object to the idea that being afraid in the jungle when hearing

---

[16] I would like to thank XXXX for raising this objection.

a specific sound is a false alarm is because it "makes sense". One can argue that one is justified and rational to be afraid in this situation. I agree, but one should not confuse "justification" with "accuracy". A justified belief might still be false: a misrepresentation can still be justified. This paper is not about the rationality of emotion, but rather about the nature of emotional representation. Cases of justified fear of things that are not dangerous are misrepresentations or false alarms.

The fourth objection also targets the definition of "danger" and could be seen as a broader criticism of the applicability of a Dretskean theory to fear. Suppose that fear does not track "danger" as such, but "response-dependent danger" (for discussion, see Brady 2016, pp. 109-117). The property which fear tracks, on this account, partly depends on the experience of fear. In other words, the property of "response-dependent danger" would not exist if there were no experiences of fear. The instantiation of the formal object depends on the experience of the corresponding emotion, according to a response-dependence theory of the formal objects. Such a theory would also define the formal objects in terms of the emotional experiences it produces. However, talking about false alarms and inaccurate representations of response-dependent danger might lead to worries about circularity. This, however, would be a problem if you would claim that phenomenology is to be reduced to content. If the experience can be reduced to its content, and the content is partly determined by the experience, then one might have to worry about circularity. However, as I do not defend this claim, as I explained in Section II, this problem does not arise for my account.

A good case could be made that fear does not track response-dependent properties. I did state in Section II that emotions represent relational properties. However, this does not imply that

they are necessarily response-dependent in the sense that the property the emotional representation tracks is not necessarily dependent on an experience. The property of danger in particular could be seen as a relational property in the sense that it is a "functional attributive" as Burge (2010, pp. 323-325) calls it.[17] The property of danger serves a function: it indicates a threat for an organism's well-being. Something is only dangerous in relation to a particular organism or person. It depends on your vulnerabilities, goals and needs whether something is dangerous. This is not necessarily connected to experience. There still is a fact of the matter whether something is dangerous, given one's vulnerabilities. In this sense, fear tracks "danger" as a relational property and it can misrepresent danger.

Even if it is the case that fear would represent a response-dependent property, there are ways to deal with worries of circularity. This debate would be similar to discussions of colour perception (see f.i. Byrne and Hilbert, 1997, 2011).[18] I do not want to commit to this theory of colour properties or to the claim that danger is a response-dependent property. I just wanted to point out that there is a way in which, if you wanted to define danger in this way, one can still agree with the claims I make in this paper.

One can also argue that fear does not represent danger or "response-dependent danger", but instead it represents the response-dependent property of the "fearsome" or "frightening" (see Tappolet, 2016). On this account the fearsome is defined as fitting fear; and fear is defined as

---

[17] Schroeter (2006) also provides a convincing defence of the claim that what fear tracks is functional and does not track something which is dependent on the experience of fear.

[18] Byrne and Hilbert define colour categories as properties that have the power to produce certain responses of sensory systems. Colour representations track these properties. Or, one might say, they reliably covary with these properties. These properties can be misrepresented. This is a way to reconcile the ideas that colour properties are dependent on sensory systems and that colour representations reliably covary with them, while avoiding circularity objections.

the fitting response to the fearsome. This theory of emotional content does seem problematically circular to me. One can argue that it is still an informative account. I would, however, disagree. At least, I believe that claiming that fear represents the functional attributive "danger", as I just explained, is much more informative than claiming that fear represents the fearsome. The "representation of danger" describes the function of fear in greater detail.

However, if you believe that fear presents us with the "fearsome", there is still a way my account is compatible with this claim. One can say that my account describes how emotions track "danger" while there is an aspect of emotion which phenomenologically presents us with "the fearsome". As I isolate my claims about emotional representation as reliable covariation from claims about emotional phenomenology, the claim that part of emotional phenomenology is "the fearsome" is consistent with the claim that instances of fear Jamesian-reliably covary with danger and thus represent danger.[19]

The fifth and last objection I will mention requires more elaboration. Suppose that John is all by himself in a cabin in the woods. He is trying to sleep but suddenly he hears a sound he cannot identify. He is scared: "better scared than sorry!". However, he remains scared all night and does not get any sleep.[20]

Compare this to Ruth. Like John, she is all by herself in a cabin in the woods. She hears the exact same sound she cannot identify. Yet, Ruth thinks: "There is no reason to panic, because

---

[19] For an argument in favor of the disentanglement between representational content and phenomenology as applied to colour, see (Shoemaker, 1994).

[20] I would like to thank XXXX for raising this objection.

I do not know what it is, so I do not have sufficient evidence to claim anything whatsoever about this sound". Ruth gets a good night's sleep.

The better scared than sorry strategy of John was not that helpful: he did not get any sleep! Ruth's Cartesian strategy seemed to be better. If we always act according to the better scared than sorry strategy, we end up being scared all the time. The pragmatic account of emotional representation is based on the idea that Jamesian reliability is pragmatically valuable for us. The sleeping in the cabin example, however, seems to indicate that Jamesian-reliable representations are not beneficial for us at all.

I hold that within a representational system where there is a trade-off between Jamesian and Cartesian reliability, it is often useful that instead of either a more Cartesian or a more Jamesian representation, one has a representation which is on the Jamesian end of the spectrum, followed by a representation which is more on the Cartesian end of the spectrum. In certain situations, it is better to have a Jamesian-reliable representation, followed by a Cartesian re-evaluation.

When interacting with a Jamesian-reliable emotional representation, higher cognition often plays the role of the Cartesian re-evaluation.[21] When you are watching a 3D movie and a snake

---

[21] This distinction between "initial response" and "re-evaluation" resembles other distinctions made in emotion theory. Some emotion scientists propose a dual system theory, consisting of two "systems": a quick and dirty system I, followed by a slow and accurate system II (LeDoux, 1998, pp. 163–164). Lazarus (1991) distinguishes between primary and secondary appraisals. The primary appraisal would represent the snake as dangerous, while the secondary appraisals would reappraise the situation in the light of new information and prevent the organism acting as if it were in a full-blown state of fear. Scherer's model includes a number of appraisals, ranging from very primitive and direct to more complex and possibly conceptual (Scherer, 2001, 2005). Robinson (2005) distinguishes between non-cognitive appraisal and cognitive monitoring in an emotional episode. In this paper, however, I do not wish to focus on mechanistic explanations of mental phenomena. I am not necessarily committed to a dual system theory. The distinction between Jamesian and Cartesian reliability might resemble

jumps towards you, the Jamesian emotion represents the 3D snake as dangerous. Cognitive deliberation, which plays the role of Cartesian re-evaluation, will rapidly calibrate this emotional representation because other evidence shows that you are watching a movie and that the snake is not a real snake.[22] In the cabin in the woods case, ideally, one would be scared at first ("better scared than sorry!"). This initial response should be followed by a Cartesian re-evaluation of the facts because one would not want to lose a whole night of sleep over it.

The cabin in the woods example was supposed to indicate that Jamesian-reliable representations are often not beneficial for us. This is in contradiction with the pragmatic theory of emotion. My response to this last objection is that it is often beneficial to have a Jamesian-reliable representation followed by a Cartesian-reliable re-evaluation.

After this final remark, I should conclude. In this paper, I have tried to solve the emotional unreliability problem: emotional representations seem to provide us with information about the world, but some emotional representations seem to be unreliable. This is a problem for the reliable covariation theory of mental representation. I have defended the pragmatic account of emotional representation. This is the claim that some emotional representations are reliable in the Jamesian sense, as they are aimed at maximizing accurate representations about relevant phenomena. They play an important role in our representational system because one is often better scared than sorry.

---

other distinctions in emotion theory but one should be careful if one wants to "map" this distinction onto other distinctions. For a detailed argument against the mapping of different dichotomies onto each other, see (Moors, 2014). Likewise, my claims about the Jamesian-Cartesian trade-off do not imply that Jamesian reliable representations are modular (see also Jones, 2006).

[22] When the 3D snake jumps towards you, you already know that you are watching a movie. The emotional representation, however, does not use this information to determine whether the object that is seemingly moving towards you is dangerous. Cartesian re-evaluation does take this information into account and because it takes time to deploy this information, it is slower.

As I mentioned in the introduction of the paper, the emotional unreliability problem is related to a larger criticism of Dretskean representationalism, namely that it cannot account for states which produce a lot of false alarms. This paper offered a contribution to this debate. On the other hand, the ideas of this paper will also help to clarify the nature of emotions. I have focused on what makes emotions similar to other mental states, so the question that remains is: how is emotion different from other mental states? As I have suggested earlier: the particular relation between emotional representation and phenomenology might be the answer. The ideas of this paper will also have consequences for the epistemology of emotion. A question one might ask is whether Jamesian-reliable emotions can epistemically justify evaluative beliefs. These are all topics that could be explored in future research.

REFERENCES

Barlassina, L., & Newen, A. (2014). The role of bodily perception in emotion: In defense of an impure somatic theory. *Philosophy and Phenomenological Research*, *89*(3), 637-678.

Burge, T. (2010). *Origins of objectivity*. Oxford: Oxford University Press.

Byrne, A., & Hilbert, D. (1997). Colors and reflectances. In A. Byrne & D. Hilbert (eds.), *Readings on Color, Volume 1: The Philosophy of Color* (263–88). Cambridge, MA: MIT Press.

Byrne, A., & Hilbert, D. (2011). Are colors secondary qualities? In L. Nolan (ed.), *Primary and Secondary Qualities: The Historical and Ongoing Debate* (339–361). Oxford: Oxford University Press.

D'Arms, J., & Jacobson, D. (2000). The moralistic fallacy: on the 'appropriateness' of emotions. *Philosophical and Phenomenological Research*, *61*(1), 65-90.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853-869.

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

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Deonna, J., & Teroni, F. (2012). *The emotions: A philosophical introduction*. New York: Routledge.

Descartes, R. (1724). *Les meditations metaphysiques*. Paris: Pierre Prault.

Döring, S. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica, 61*(3), 363-394.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content and function* (17-36). Oxford: Oxford University Press.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Field, H. (1990). "Narrow" aspects of intentionality and the information-theoretic approach to content. In Villanueva, E. (Ed.), *Information, semantics, and epistemology* (102-116). Oxford: Blackwell.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Godfrey-Smith, P. (1991). Signal, decision, action. *The Journal of Philosophy, 88*(12), 709–722.

Godfrey-Smith, P. (1998). *Complexity and the function of mind in nature*. Cambridge: Cambridge University Press.

Goldie, P. (2007). Seeing what is the kind thing to do: Perception and emotion in morality. *Dialectica, 61*, 347–361.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Gordon, R. (1987). *The structure of emotions*. Cambridge, UK: Cambridge University Press.

Greenspan, P. S. (2006). Practical reasoning and emotion. In A. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (206-221). New York: Oxford University Press.

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*(3), 390–421.

James, W. (1899). *The will to believe and other essays in popular philosophy*. New York: Longmans, Green, and Co.

Jones, K. (2006). Quick and smart? Modularity and the pro-emotion consensus. *Canadian Journal of Philosophy, 36(*1), 2-27.

Kenny, A. (1963). *Emotion, action and will.* London*:* Routledge and Kegan Paul.

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

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(8), 2871–2878.

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science*, *25*(9), 1663–1673.

Macmillan, N. A., & Creelman, C. (2004). *Detection theory: A user's guide*. Hove: Psychology Press.

Millikan, R. (1989). Biosemantics. *The Journal of Philosophy, 86*, 281 297.

Millikan, R. (1990). Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics. *Philosophical Topics, 18*, 151 161.

Mitchell, J. (2019). Affective Representation and Affective Attitudes. *Synthese* (4), 1-28.

Moors, A. (2014). Examining the mapping problem in dual process models. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (20–34). New York: Guilford.

Moors, A. (2017). Integration of two sceptical emotion theories: Dimensional appraisal theories and Russell's psychological construction theory. *Psychological Inquiry, 28*(1), 1-19.

Müller, J. M. (2019). *The World-Directedness of Emotional Feeling: On Affect and Intentionality*. Cham, Schweiz: Palgrave Macmillan.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions.* Cambridge: Cambridge University Press.

Papineau, D. (1987). *Reality and representation*. Oxford: Blackwell.

Price, C. (2015). *Emotion*. Cambridge: Polity.

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

Quine, W. V. O. (1969). *Ontological relativity and other essays.* New York: Columbia University Press.

Robinson, J. (2005). *Deeper than Reason: Emotion and its role in literature, music, and art.* Oxford: Oxford University Press.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of personality and social psychology, 50*(4), 703-712.

Scarantino, A. (2010). Insights and blindspots of the cognitivist theory of emotions. *British Journal for the Philosophy of Science, 61*(4), 729-768.

Scarantino, A. (2014). The motivational theory of emotions. In D. Jacobson & J. D'Arms (Eds.), *Moral psychology and human agency* (156-185). Oxford: Oxford University Press.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (92-120). New York: Oxford University Press

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information. Information Sur Les Sciences Sociales, 44*(4), 695–729.

Schroeter, F. (2006). The Limits of Sentimentalism. *Ethics, 116*(2), 337-361.

Shoemaker, S. (1994). Phenomenal character. *Noûs, 28*(1), 21-38.

Smith, C. A., & Kirby, D. (2009). Goals. In D. Sander & K. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (197-198). Oxford: Oxford University Press.

Solomon, R. (2003). What is a "cognitive theory of the emotions", and does it neglect affectivity? In *Not Passion's Slave* (178-194). Oxford: Oxford University Press.

Stich, S. P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Swets, J. A. Green, D. M. (1964). *Signal detection and recognition in human observers: Contemporary readings*. New York: Wiley.

Tanesini, A. (2008). Virtues, emotions and fallibilism. In: Brun, G., Doguoglu, U. and Kuenzle, D. (Eds.), *Epistemology and the Emotions* (67-82). Aldershot: Ashgate.

Tappolet, C. (2016). *Emotions, value, and agency*. Oxford: Oxford University Press.