Jones, Karen. 2012. Response to Nancy Daukas.
Social Epistemology Review and Reply Collective 1 (10): 1-7.
http://wp.me/p1Bfg0-sN

Response to Nancy Daukas

Karen Jones, University of Melbourne

When I first began to think about the emotions, I thought I knew what belief was. Belief was the easy side of the belief/emotion contrast, something that could be taken as given while doing the hard job of working out what emotions are and the complex ways in which they relate, and sometime fail to relate, to belief. Not surprisingly, things turned out not to be so simple: the more I thought about the emotions, the more I felt my grip on belief weakening. This much I do know for sure: the mind is messier than philosophers' models tend to acknowledge. According to one such influential model, brought to prominence by Davidson (1963), the contents of the mind divide into two halves according to direction of fit. The belief half is characterized by a mind-to-world direction of fit, so that beliefs tend to go out of existence with awareness that the world is not as the belief represents it to be. The desire half is characterized by the reverse world-to-mind direction of fit, so that desires go tend to go out of existence when the world is judged to be as desired. Together, belief and desire explain action. The desiderative half is taken to include emotions, preferences, and any kind of "pro attitude", thereby giving us a cognitive/desiderative divide, with emotions being on the non-cognitive side. Thinking seriously about emotions problematizes this classification, for emotions are, in de Sousa's (1987) apt phrase, "Janus-faced," simultaneously representing the world as being in a certain way and as to be made to be in a certain way, but at the same time resisting decomposition into beliefs and desires. Emotions belong to a messy territory in between the traditional cognitive and non-cognitive divide and so, too, I argue does trust, whether practical or theoretical and whether trust in oneself or in others.

Nancy Daukas's discussion of my paper, "The Politics of Intellectual Self-Trust," rightly calls for further investigation of the relationship between belief and affect. She is correct to worry that in moving from the fact that our habits of intellectual self-trust respond only slowly and imperfectly to changes in our judgment about our reliability (in a domain) to the claim that self-trust is best understood as a domain-relative attitude of optimism about our cognitive competence, I have overlooked an alternative explanation. Perhaps those whose patterns in their intellectual self-trust depart from what, on reflection, they assert about their own competence harbor deep beliefs that are in opposition to their assertions. In cases where self-trust has been molded by racism and sexism, perhaps, deep down, they harbor entrenched racist and sexist beliefs, beliefs that they would be unwilling to avow not only for the external reason of fear of disesteem but also because such beliefs are incompatible both with their self-conceptions and with what they take the evidence to be. This alternative belief-based explanation, Daukas presses, might even be better able to explain the dispositions that I have claimed are constitutive of intellectual self-trust than my alternative affective analysis. Consider, for example, the disposition confidently to assert what you believe in domains in which you have self-trust, a disposition that might be taken to express a belief in your reliability.

Daukas and I are in agreement about the practical conclusion of my argument: pathologies in intellectual self-trust, whether of excess or deficit, and especially those pathologies that are generated by social relations of dominance and subordination, are not

easy to fix. We cannot get our house in order as inquirers simply by accepting, on an intellectual level, the evidence that we have excesses or deficiencies here. We have to go through a slow process of rehabituation to bring our day-to-day responses into alignment with the theories we accept. Lining up all your prizes up in a row and running inductions about your past achievement is not going to help you if you are in the grip of a socially generated acute deficit of intellectual self-trust. (Anecdotal evidence from talking to many feminist philosophers about crisis moments in their careers — and I know of none who have entirely escaped crisis — strongly supports this claim.) Likewise, accepting, however sincerely, the now well-supported theories about how implicit bias works is not going to cure you of it. My rejection of belief-analyses of intellectual of self-trust is driven, in part, by the hunch that they prescribe overly shallow remedies for its pathologies; but as Daukas points out, why suppose that there is anything shallow about belief? Why, then, did the inference from the fact that habits of intellectual self-trust seem sometimes to be in tension with judgments about reliability and respond to such judgments only slowly and often imperfectly to the view that self-trust is affective seem compelling to me? Behind that inference lie background assumptions that I have taken directly from work on the emotions where this kind of mismatch between judgment and affect – called "recalcitrance" – has played a central role in belief accounts of the emotions falling from favor. What I want to do in this rejoinder is unbury some of those assumptions and offer more by way of explication and defense of them, as well as further support for their applicability to the case of intellectual self-trust.

In developing my account of intellectual self-trust, I start from pathological cases, such as the neurotic passport checker and the pigheadedly self-confident person. Daukas wonders whether this is the right place to begin, given that pathological cases do not follow normal patterns. Typically, they are both statistically abnormal and also departures from norms governing how our responses ought to be (though socially engineered pathologies can entrench so as to become statistically normal, all the better to fulfill their role in maintaining relations of dominance and subordination). Might we distort our understanding by starting here, with the pathological, rather than starting with the normal case? There's a reason to start where things are going wrong rather than with where they are going right: breakdowns isolate moving parts, parts that we can fail to notice when things are working smoothly. For this reason, many contemporary discussions of the emotions start with recalcitrance. Recalcitrance exists on a continuum from the clearly pathological, such as phobias, to the everyday phenomenon of emotional lag, as when our anger lingers on even after we realize it rested on a mistake. Recalcitrance is exhibited when our feelings fail to fall in line with our judgment about the evaluative properties of a situation. Because emotions are recalcitrant they can provide us with an important corrective to poor evaluative judgment, which they could not do if they moved in lock-step with judgment. Our emotional responses are possible evidence regarding evaluative features that our judgment might have overlooked. This means that there are pluses as well as minuses to recalcitrance. The phenomenon of recalcitrance suggests that our emotions are a second system of evaluation that is at best only partly integrated with the reflective system (Tooby and Cosmides 1990, Prinz 2004, Jones 2003). Moreover, this might not be a bad thing – emotions might contribute positively to human rationality by providing fast answers to important practical problems.

Thinking about recalcitrance led many to abandon analyses of emotions that see them as fully or partly constituted by beliefs. The path to rejecting belief accounts is not, however, as simple as noting that emotions are a separate evaluative system that enjoys only partial integration with our reflective system; for even granting this is so, the output of the affective system might yet be an evaluative *belief*. A further argument is needed to show that the output is not itself a belief (though it might, and frequently does, give rise to a belief). That argument is provided by detailed examination of cases of phobias. Consider Patricia Greenspan's case of the person whose past traumatic experience with a dog makes her afraid of all dogs (Greenspan 1988, 17-20). She is afraid of toothless, old, arthritic Fido, even though she judges that Fido is not dangerous. To judge that Fido is not dangerous is not yet to believe that he isn't: a judgment is an occurrent intellective act of assent to a proposition; beliefs, on the other hand, are multi-stranded dispositions. (So I should not have slid so casually between judgment and belief, as I did in the earlier presentation of my argument.) Our judgment that *p* is not guaranteed to set in place the complex suite of dispositions that constitutes believing that *p*, or to displace the dispositions constitutive of believing that not-p once they are firmly lodged in our heads. Perhaps — though this is disputable — an ideally rational agent would move without friction from judgment to belief, but we normal agents are not frictionless and can fail to meet this ideal.

To show that belief need not accompany emotion we need to carefully examine the dispositions had by someone with a phobic fear of Fido and those had by someone who believes that Fido is dangerous. According to the standard analysis of belief, beliefs are states that play a certain functional role. At the highest level of description, they are states with a mind-world direction of fit and thus believers must have some, albeit imperfect, disposition to modify their beliefs in the light of their understanding of the way the world is. Beliefs also combine with other beliefs in inference to generate further beliefs and join with desires to explain action. Thus, to believe that Fido is dangerous is, *inter alia*, to be disposed to make inferences on the assumption that Fido is dangerous, to be willing to assert that Fido is dangerous, to take measures to protect oneself and others from Fido, and so on. The specific profile that a belief that Fido is dangerous displays in any given context is a function of the believer's other mental states, including both beliefs and desires. Someone with a phobic fear of toothless, arthritic, old Fido typically lacks enough of this profile to count as believing that Fido is dangerous. Although there is some overlap between her suite of dispositions, such as Fido-avoidance, she is not disposed to warn others to stay clear of Fido, to assert that Fido is dangerous, to take measures to protect herself from Fido and so on. The phobic lacks too many of the dispositional strands of those who believe that Fido is dangerous to count a believer in Fido's dangerousness.

Perhaps it might be objected that the strands the phobic lacks are only those of someone who is willing to avow that Fido is dangerous and so someone who has a conscious belief in Fido's dangerousness. Unconscious beliefs might be though to operate differently, including with respect to how they are regulated by the agent's understanding of evidence. On this alternative picture, the mixed set of dispositions that the phobic has is explained by her having a conscious belief that Fido is not dangerous (hence the

behavioral overlap with those who hold this belief in assertion, dispositions to warn, etc) and the unconscious belief that Fido is dangerous (hence her departure from the profile of someone who believes that Fido is not dangerous, as a portion of those constitutive dispositions are knocked out by opposing ones).

But this alternative picture won't do, either. Charity counts against it (Greenspan 1988, 19-20). It assimilates the irrationality of the phobic to the irrationality of both believing that *p* and believing that not-*p*; but the phobic's problem doesn't seem to be of that obvious kind. We need grounds independent of the fact that the phobic is afraid of Fido before we ascribe inconsistent beliefs to her. But insofar as she lacks key elements of the functional role of believing that Fido is dangerous we lack such independent grounds.

It is not as if the phobic's fear is without cognitive content, however. Rather, the way it functions cognitively resists being forced into the shape determined by a functional role analysis of belief, suggesting that the traditional cognitive/non-cognitive divide in which belief takes up the whole domain of the cognitive is too crude. There is much more on the cognitive side than belief. There are a variety of accounts of how best to characterize the cognitive component of emotions, but what is common ground is that the functional role emotions play in cognition centrally includes shaping attention, directing thought, and influencing inference. Emotions determine salience and, because of this, they readily influence belief, desire, and, hence, action. Their nearest analogy is to perceptions.[1] Many theorists think of them as evaluative appearances. Like optical illusions, such as a straight stick appearing bent when half submerged in water, affective appearances resist correction in the light of things we believe. At most we can insulate them from affecting what we go on to believe, but if we are not careful they will give rise to beliefs. Visual and emotional systems are, though to different degrees, informationally encapsulated.

In making the inference from the recalcitrance of self-trust to an affective account of it, I was tacitly calling on this account of the emotions and assuming that the pathologically self-distrusting person is strongly analogous to the phobic. Less pathological cases, I assumed, were to be located at various mid points along the recalcitrance spectrum, where there would be even less reason to ascribe contradictory beliefs. But does the analogy hold? And if it does, what does this say about the cognitive content of intellectual self-trust? Daukas is concerned that my account doesn't locate that cognitive content clearly enough. On the one hand, she presses, if self-(dis)trust is simply a matter of feelings of self-confidence, detached from cognitive content, then my account isn't a mixed cognitive/affective one at all. On the other hand, if the dispositions the self-(dis)trusting display are not expressions of beliefs or beliefs overlaid with or accompanied by such optimistic feelings, then what exactly are they expressions of? Where is the cognitive element?

[1] Perceptual theories are the dominant contemporary approach in philosophy of emotions. Key proponents include de Sousa (1987) and Prinz (2004). There is significant variation among perceptual theories. For an overview and further references see Deonna and Teroni (2012).

Let me first clarify something: I am not meaning to claim that the self-(dis)trusting never believe that they are(not) reliable in a domain. These beliefs do indeed tend to accompany self-(dis)trust. I only mean to claim that they do not invariably accompany it. Belief and attitude can part company, on the model of recalcitrant emotions; thus, our analysis should not make so-believing constitutive of self-(dis)trust. Whether it correct to interpret someone as believing that he or she is (un)reliable is going to be a matter of that person's total cognitive, desiderative, and behavioral profile.

Applying these background assumptions about emotions to the case of intellectual self-trust gives us the following account of the cognitive content of intellectual self-trust. In the first instance, those with intellectual self-(dis) trust, in a domain, experience an emotionally laden perception of the situation. This perception has cognitive content: the situation is experienced as cognitively safe (risky). This perception typically has a distinct phenomenology and sets in train the cognitive functional roles characteristic of affect: to control salience, direct thought, and influence inference. Take the case of the self-distrusting. When deciding what to believe in a domain where they lack intellectual self-trust, the self-distrusting experience the situation as presenting cognitive risk. The possibility of error is salient to them and this salience is, I claim, enough to explain the dispositions characteristic of the intellectually self-distrusting. These dispositions are the opposite of those characteristic of self-trust, and so include include, in the relevant domain, being disposed to feelings of lack of self-confidence, hesitancy to assert, discounting one's own judgment especially in the light of conflict, and rumination on one's competence. (There is room for various shades of neutrality in between as trust and distrust are mutually exclusive, but not exhaustive.) So long as the possibility of error is at the forefront of their minds, it is plausible that they will have all these dispositions. We need not also ascribe to them a belief in their incompetence to explain these dispositions. Whether it is appropriate to ascribe this belief is going to depend on the whole of their cognitive profile, which will sometimes support ascribing the belief and sometimes not.

I not only claimed that it is possible to believe that you are reliable in a domain and yet have the salience patterns and, hence dispositions, characteristic of the self-distrusting but also that it is possible to believe you are unreliable and yet experience situations as presenting little cognitive risk and so have the salience patterns and hence dispositions characteristic of the self-trusting. This claim might be thought less plausible. Surely if you *really* believed you lacked competence in a domain, you would not think, infer, and act as you do. You must either not really believe it, or have some "residual" opposing belief that's preventing your belief from playing its usually functional role. But this seems to me incorrect, also. The distinctive cognitive role of emotions is to control salience: emotions do this, beliefs do not. Consider the use of graphic packaging images as part of anti-smoking campaigns. Smokers believe, indeed know, that smoking is dangerous to their health. But this belief is not at the forefront of their minds as they reach for another cigarette. The images are designed to induce fear so as to make thoughts of danger inescapable in the hope of thereby modifying behavior despite the opposing force of addiction. "Hot" or affective cognition recruits things believed and sometimes only imaged and brings them to bear on the problem at hand.

None of this is to say people who believe they are not reliable would inevitably fail in their attempts to make that belief get purchase in practice unless it were accompanied by an affective attitude of pessimism about their competence. If they are neither optimistic nor pessimistic and approach with a stance of neutrality, the chances of that belief getting purchase must surely increase. But even if they continue to experience situations as providing little cognitive risk — as I take it many who continue, case by case, to be confident in their judgment despite accepting the work on implicit bias do — there is still the chance of cognitive override. The world is experienced, in affect, as being a certain way but to so experience it is not yet to believe that the world is as it presents itself as being. On my account, there remains a gap between presentation and belief. This gap is a potential space for self-regulation, but the speed and pre-reflective nature of emotional processing suggests that such self-regulation will be difficult.

Daukas is more optimistic about the role of meta-cognition as a force for change than I am. I'm willing to concede that, driven by the sense of how very hard correction is and how deep the problem goes, I might have over-stated the case for pessimism. But pessimism does have empirical support in recent work in psychology on fast or System 1 cognition that is automatic, intuitive, typically affective, and biased in various ways (Kahneman 2011). System 1 provides "quick and dirty" information processing. From studies of risk assessment, Paul Slovic et al (2002) posit an affect heuristic, arguing that how you feel about a technology — that is, your emotional response to it — affects your assessment of its merits. Kahneman (2011, 97-105) argues that the affect heuristic is an instance of question substitution: when, regarding some complex matter, we are confronted with the difficult question "what do I think about it," we substitute instead the easier to answer question, "how do I feel about it." We do not notice this substitution and so we think we have answered the original target question. Applying this model to the case of meta-reflection regarding whether we are trustworthy in a domain grounds pessimism. Unless we are really careful, we will be likely to side-step proper deliberative confrontation with the right question, "am I really reliable, here?" and substitute instead the question of how we feel about our own competence. If we do that, the degree of intellectual self-trust with which we began will simply be confirmed.

A natural question for anyone who has proposed an account of intellectual self-trust is, "do you deserve your own trust in this domain?" Fortunately, by thinking together, we increase our trustworthiness and that has happened in this exchange. I was too quick to dismiss accounts that insist intellectual self-trust must contain belief, even if belief alone is not the whole story. I hope I have done something to rectify that here. When approaching problems in practical epistemology — and I take it the problem of how to rectify excesses and deficits of self-trust is one such problem — the thinking together that must be done is necessarily interdisciplinary. Philosophers have much to learn from psychologists, but I don't doubt that just as philosophical accounts of the mind have too been too simple, current psychological models are too simple also. Many of those models imply that it would take a kind of super-human effort of cognitive override to eliminate intuitive bias, but emerging evidence suggests our reflective capacities, when combined with enough motivation to change, can over time transform our fast intuitive processing

(Monteith et al, 2009). Clearly, though, to trust ourselves now in these domains would be unwise.

Contact details: jonek@unimelb.edu.au

References

Davidson, Donald. 1963. Actions, reasons and causes. *Journal of Philosophy* 60: 685-700.

Deonna, Julien and Fabrice Teroni. 2012. *The emotions. A philosophical introduction*. New York: Routledge.

Daukas, Nancy. 2012. Comments on Karen Jones, 'The Politics of Intellectual Self-Trust'. *Social Epistemology Review and Reply Collective* 1 (6).

de Sousa, Ronald. 1987. *The rationality of emotion*. Cambridge: MIT Press.

Ekman, Paul. 2003. *The face revealed: Recognizing faces and feelings to improve communication and emotional life.* New York, NY: Times Books.

Greenspan, Patricia. 1988. *Emotions and reasons: An inquiry into emotional justification*. New York: Routledge.

Jones, Karen. 2003. Emotion, weakness of the will and the normative conception of agency. In *Philosophy and the emotions,* edited by Anthony Hatzimoysis, pp.181-200. Cambridge University Press.

Jones, Karen. 2012. The politics of intellectual self-trust. *Social Epistemology* 26 (2): 237-251.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Montheith, Margo, Jill Lybarger and Anna Woodcock. 2009. Schooling the cognitive monster: The role of motivation in the regulation and control of prejudice. *Social and Personality Psychology Compass* 3 (3): 211-226.

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

Tooby, John and Leda Cosmides. 1990. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11: 375-424.

Slovic, Paul, Melissa Finucane, Ellen Peters, and Donald G. MacGregor. 2002. The affect heuristic. In *Heuristics and biases,* edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, pp. 397- 420. New York: Cambridge University Press.