

# Research Plan: In Defence of Affective Benevolence

*Principal Investigator:* Prof. Julien Deonna, University of Geneva, Department of Philosophy.

## 1. Current state of research in the field

Benevolence is a good thing. While it need not be subtended by emotions, it often is. Compassion, love, joy, gratitude, but also guilt and indignation, play an important role in the fact that we behave in caring ways. Emotions are important not only in the sense that they are frequent determinants of caring behaviour. Because of their intimate connexion with felt motivation, emotions arguably lend special worth to the practices of giving and helping. There is something very natural to the thought that it is good to be moved to help by compassion, to want to repair our faults through guilt, to want to change the world for the better through our indignation, or simply to want to benefit others because we love them. This project is about the role and status of such emotions in benevolence, that is, actions done for the sake of other people. Now, although such emotions do indeed play these roles and on the whole have enjoyed this good press, philosophers have always had their doubts about whether emotions can be considered a force for good when addressing the question of our benevolent inclinations. This unease has recently been expressed in a particularly vehement form by both philosophers and psychologists and has had important repercussions in the field of philanthropy.

There are at least two routes to the distrust against emotions in this context, what we will call (1) “the *cynical* route” and (2) “the *sceptical* route.” The gist of the project aims quite generally to develop the argument (against the cynic) that emotions play an ineliminable role in grounding our understanding of why the world needs us to be benevolent, and, more generally (against the sceptic), in constituting our identities in the cultivation of the attachments to people that need our care, as well as to impersonal values of beneficence that claim that we honour them. Additionally, the project aims to contribute to the ethics of benevolence by exploring how these ideas play out within the field of organized philanthropy.

(1) *The cynical route.* One source of the distrust we may have with regard to emotions in relation to benevolence is as old as philosophy itself: while emotions subtending benevolent action may look like the manifestations of our concern for others, they would in fact only betray our concern for ourselves. That is the picture of human action Hobbes canvasses in his *Leviathan* (1651/1991) and finds incisive expression in the satirical writings of Mandeville’s *The Fable of Bees* (1714/1997). In the contemporary philanthropic literature, this tradition is pursued with much vigour and ingenuity by economist James Andreoni (1990), and many others who claim that the institution of giving can only make sense in light of the internal rewards we reap from said activity, namely, the “warm glow” we feel as we witness the effect of our benevolence. Let us call this “the cynical route” to the misgivings we may have vis-à-vis emotions. In the case of compassion or guilt, say, people are only concerned about getting rid of the unpleasant feeling connected to these emotions. The exact motivations behind these claims vary, but the dominant idea is the thought that all human pursuits must ultimately rest on self-interest. We give for the satisfaction provided by giving. We help for relieving the frustration we feel witnessing other people’s distress, etc. This powerful explanation of human behaviour is easy to understand, and teachers of philosophy

everywhere have no difficulty conveying its appeal to their students. If generosity and compassion seem to manifest concern for others, says the cynic, only a superficial understanding of these emotions may lead one to believe in our good nature.

While not easy to rebuke, good answers to the cynical challenge have existed for a long time and their expression in 18<sup>th</sup> century sentimentalism have been particularly sophisticated (Hutcheson, 1725/1991; Hume, 1751/1998; Smith, 1759/2010). Some of these answers have been *empirical* in nature. If we are not convinced that the efforts, big or small, that people endure for the sake of others are a sign of our altruistic nature, it should be obvious that the egoist explanation of the major sacrifices we are capable of is not persuasive. For example, people could easily choose other means than helping to rid themselves of the pain they feel at the distress of others. If it were true that people wanted to benefit themselves first and foremost, they would find easier means than the pursuit of benevolent action to reach this goal. But they often do not. It is precisely this type of empirical evidence, presented in anecdotal form by the sentimentalists in their days, which was pursued with great systematicity in some of the most interesting work of late 20<sup>th</sup> century moral psychology. Batson and his collaborators, in research spanning over more than three decades, set out to devise experimental ways to demonstrate that subjects did indeed behave in altruistic ways (Batson, 2011). Overall, findings emanating from a very varied set of experimental settings consistently show that subjects go for the altruistic route even when egoistic alternatives are available at little costs to themselves.

Other answers have been *conceptual*. Perhaps the most influential argument in this area, already traceable to the writings of Butler (1726/1991) and Sidgwick (1907/1981), has it that, even if to get satisfaction and avoid frustration were all that people ever pursued in life, the very pursuits in question do presuppose that people care about things other than bare satisfaction-seeking and frustration avoidance. People sometimes care about the satisfaction that comes with benefitting others and *this* is very different from the satisfaction associated with other goals they might have in connexion to more personal goals. And if people do care about this kind of satisfaction, then they do care about good things happening to individuals other than themselves (Feinberg, 1965/1999; Broad, 1930/2000). Similar-sounding arguments have been made directly about the emotions, e.g., pleasure, compassion (Tappolet, 2000), or guilt. Regarding the latter, for example, there is no shortcuts to its elimination, at least in normal circumstances. We would not want to take a magic pill to get rid of the guilt we feel even if such pills existed and were easily available. If we do feel guilt, it is usually because we think we have done harm and, typically, the only means to getting rid of that feeling that does amount to postponing the problem is to repair or compensate the harm (Elster 2006). Once again, it is not simply the alleviation of discomfort that we seek, but alleviation of the right kind, specific for guilt (and, presumably, different kinds of alleviations may be specific to other emotions like compassion, etc.).

It is fair to say that a majority of philosophers in the area claim to be convinced by these arguments. Many are not, however. Regarding the empirical evidence, it is often pointed out that there often remain plausible egoistic explanations as of why people pursue altruistic behaviour. Yes, subjects help when they could easily avoid it, and yet that is because, say, they fear for the tranquillity of their sleep were they to fail to do the right thing. Their motives, that is to say, remain self-centred. Despite the ingenuity of the protocols in the relevant experiments, the hardened cynic always finds a story to accommodate her hypothesis (e.g. Stich, Doris, & Roedder, 2010 for

discussion). The same goes for the conceptual arguments. The cynic needs not deny that subjects are indeed moved by their concerns for others; what they have to deny is that these concerns are the ones which *ultimately* explains why subjects do act in the way they do (Sober & Wilson, 1998; May, 2011; Clavien, 2012). And they can argue that nothing in the arguments as they stand succeeds in showing that concern for others is not just a means, even if perhaps the only one available, to an ultimate self-serving aim, that of getting an emotional kick or simply recovering much needed emotional balance (Mercer, 2001).

These arguments, both empirical and conceptual, continue to this day in very sophisticated forms (Elster, 2011; Clavien, 2012; Kraut, 2016). Yet elements of progress in the relevant debates are few and far between. The cynic seems to have endless resources in her capacity to appeal to unconscious selfish motives (Simler & Hanson, 2017). The debates, we surmise, may advance if we construe it differently, pressing novel and under-theorized questions. For one striking fact about present discussions on the topic is that, although they do focus on motives, give little to no consideration to the nature of the emotions featuring in the relevant scenarios (but see Bianchi, Cova & Tieffenbach, 2023; Tappolet, 2023, ch. 10). This project supposes that progress can be accomplished by paying attention to some plausible assumptions about the nature of emotions, that is, by situating the questions of the debate within a broader understanding of human action and its interconnexion with one's affective life. To anticipate, looking closely at the way we experience others in connexion to us while going through these emotions will make the cynic's case much more difficult to make.

(2) *The sceptical route.* The second source of unease concerning the role of emotions in benevolent action can be called "the sceptical route". We started with the seemingly innocuous observation that many of our benevolent thoughts and actions are grounded in such emotions as compassion, guilt, love, etc., where "grounding" seems to mean that these emotions cause people to have benevolent thoughts and motivate them to behave in benevolent ways. From this observation to the idea that cultivation of these emotions is a goal worth pursuing there is a small step that many have indeed taken across the ages, from Aristotle to the already mentioned sentimentalists (Hutcheson, 1725/1991; Hume, 1751/1998), to present-day virtue ethicists (e.g. Hursthouse, 1997; Roberts, 2013; Kristjánsson, 2015, 2018; Zagzebski, 2020, 2023). Now, contrary to the cynic, the sceptic does not need to deny that emotions sometimes are a positive force in our benevolent behaviour. What she denies is that we can harness these emotions to drive peoples' benevolent behaviour in any consistent, reliable or benign way. Emotions fail to be a positive force not because they never are but because they too often exert their power when they should not and too often fail to do so when we most need them. Why so? For many reasons but let us focus on four.

First, emotions are subject to *spotlight* effects. Emotions are more easily triggered by what we can directly experience or easily imagine than by abstract information. The result is that faraway people and complex problems often fail to move us when perhaps they should. Second, emotions are *biased*. Our feeling bad for the distress of others, for example, does not seem to correlate with the real distribution of distress we encounter, but with the distress of those we feel closer to. As a result, only those we are already favourably biased towards benefit from our benevolent actions, neglecting many who would profit as much or perhaps more. Third, emotions are *fickle*. Bad mood, tiredness, lack of focus, slight discomforts, fear of authority, etc., are enough to derail our "normal" emotional responses. The result is again that the benevolent treatment people receive on account of emotions seems often random. Fourth, the emotions that tend towards benevolent action are *unsafe*. Indeed, compassion

and guilt can be emotionally draining to the point that it may be the ones who feel them that find themselves in need of benevolent action. All in all, then, our sceptic argues that we should not wish for a world in which the provision in wellbeing it cries for was secured by our capacity for benevolent emotions (see Bloom, 2016; Prinz, 2011; Camassa, 2023 for a summary of the relevant empirical literature). There should be no surprise in that: the function of these emotions, like all the others, has been shaped in physical and social settings that have nothing to do with our modern-day living environments. As a result, it would be ludicrous to think that they are the tools needed to confront the global moral challenges of our age (Street 2006). That is in broad strokes what a sceptic about such emotions claims. As we shall see, the significance of the claim might seem very different depending on what the initial value we attribute to emotions in the context of benevolence was supposed to cover.

For now, it has to be noted that the reaction to this attack on the emotions over the last twenty years or so have had the following results. Simplifying a great deal, one trend has consisted in putting aside emotions entirely, while presenting consequentialism – the claim that nothing apart from the good consequences of our actions for all concerned should count in matters of moral choice – as the only enlightened and progressive ethical approach in the area (Singer, 2009). This move was made all the more attractive to some by the portrayal of alternative deontological moralities – claiming that ethical choice is a matter of doing one’s duty *for* the reason that duty requires it– as being sentimentalism in disguise. The invocation of duties, it was argued, is merely a way of rationalising our emotional impulses (Greene, 2007; Haidt, 2001). This trend has had a tremendous influence on scholars directly involved in the practical side of organised benevolence. The very popular “effective altruism” movement, which advocates using *reason* and *evidence* to figure out how to do the most good in benefiting others, has caught the attention and the imagination of many by recommending that emotions should cease to play any role in our philanthropic endeavours for precisely the reasons that they tend to favour allocations of resources that are inefficient and unjust (MacAskill, 2015, 2022; MacAskill, Mogensen, & Ord; 2018; Singer 2015).

A less radical reaction to emotions’ volatility can be found in the “rational benevolence” approach to the issues at end (Crisp, 2006, 2008; Bloom, 2016). Rational benevolence, with its emphasis on the universal, or neutral, point of view, is utilitarianism rooted in the virtues of compassion, justice, kindness and more generally the idea that welfare is not just getting what one wants. To illustrate, our concern for the welfare of all should be sensitive to such things as the fact that some are more unfortunate than others (therefore meriting more of our compassion) or that some are getting more than they deserve (meriting indignation), etc. Looking closer at the idea, however, it is unclear that the model really tabs into the emotions of any real human beings. That is explicit in Crisp, who speaks of “idealized emotions,” perhaps less so in Bloom, who describes compassion as a form of global well-whishing.

Contemporary sentimentalists have rejected the unnerved reactions of present-day utilitarian thinking on these matters, by emphasizing that the volatility of emotions is something sentimentalists have always been aware of. Therefore, they suggest that the remedy they have recommended can be refined in ways that do not require to discard our affective nature altogether. The most developed form of this response in ethics can be seen in the writings of D’Arms and Jacobson (D’Arms & Jacobson, 2023), and, in a perspective that emphasises the epistemic dimension of emotion, those of Linda Zagzebski (Zagzebski, 2020). The first thing a defence of the importance of emotion in the area must stress is the difference between the “incidental” and the “integral” influence of

emotions (e.g. Loewenstein & Lerner, 2003). True, emotions have arational influences that may conspire to produce the perverse effects described above, and yet this should not obscure the fact that emotions can and often should be viewed as rational responses to what happens around us (de Sousa, 1987). Underserved suffering, unfair distribution of public resources, the avoidable harm we inflict to others are *reasons* for, respectively, our compassion, indignation and guilt. While it is true that our emotional capacities may go astray in countless ways, it is a mistake to think that they cannot be regulated in ways that may attune them to the circumstances that triggered them. This regulation, it is believed, more often than not takes the form of our rational capacities intervening to monitor the proper operations of our emotional sensibilities (D’Arms & Jacobson, 2023). This monitoring can be deliberate and conscious as when the emotion has alerted us of something to warrant the search for rational explanations of what is happening (Brady, 2013); or it can take the shape of a more automatized monitoring, a meta-sensibility of sort (Tong, Teo, & Chia, 2014; Boisserie-Lacroix, 2023), just insuring that the circumstances would not defeat our emotional verdicts (Hutton, 2023). This second-nature type regulation of our emotions evolves alongside the development of our emotional sensibility in general and of our capacity to empathise, simulate and factor in the emotional point of view of others (Camassa, 2023): as it stands, it is an integral part of this sensibility and skills (Kauppinen, 2014a; Kristjánsson, 2018). Three interrelated points are often made and are important to emphasize in the present context. First, it would be a mistake to think that our rational capacities are insulated from these factors that variously bias, blind or disrupt our emotions (Deonna & Teroni, 2012, 2022). Being also liable to the influence of these factors, our rational capacities should not be trusted blindly. Second, and perhaps even more significantly, a lot of work in the epistemology of emotions in recent years has been focused on the idea that emotions might be shielded from certain prevalent norms or prejudices around us in ways that value judgements may not be (Arpaly, 2000; Döring, 2012; Tappolet, 2003; Silva, 2021). Although we convince ourselves through careful deliberation that some course of action is the way to go (“I should stay in the marriage for the sake of the children”), our emotions often “know better” and divert us onto another course (“Let’s make it over and leave!”). If that is the case, then, whatever precautions need to be taken before committing to an ethically significant action, these are called for whatever the sources of our behaviour’s motivations. Third, rich portrayals of how emotions ground the normative dimensions of human *relationships* through the way they contribute to mutual understanding have been an important addition to the recent scholarship on the value of emotions in ethical life (e.g. Slote, 2010, 2017; Betzler, 2019; Bailey, 2020). Contemporary sentimentalists, then, have done a lot to show that the scepticism with regard to the value of emotions in grounding benevolence tends to overreach; at the same time, they have done little to undermine their significance when they do seem to motivate in the right way.

What the state of the art makes apparent is that much of the confusion that we encounter when probing the literature on the role of emotions in ethically relevant behaviour is the exact brief against which we should assess their merit. Are we talking about the role of emotions in determining the *token* occasions in which benevolence is called for, or are we talking about the role of emotions in determining the *type* of occasions that makes benevolent actions appropriate? Are we talking about the role of emotions in *shaping* our sensibility to the demands other people’s welfare might have on us, or their role in exercising this sensibility? Are we talking about emotions’ role in *benevolent* action or are we talking about the role of emotions in *ethical* decision more generally? Are we talking only about their role in ethically relevant decisions taken by individuals in their daily ordinary lives or are

we also talking about the same decisions taken by groups or institutions entrusted with a mandate to do good? Additionally, how answers to these questions are impacted if and when we focus on the particular context of organised philanthropy? These questions are rarely distinguished and keeping them separate is, we believe, an aim the relevant literature should aspire to. This being said, our contention is that answers to all these questions are best served by focusing primarily on the role of emotions in our *understanding* (Brady, 2013; Deonna & Teroni, 2020, 2022c, Ouellette-Dubé, 2022) the demands imposed on us by other people's welfare.

The hypothesis is that, once this role become apparent, we find ourselves equipped with the means of providing a coherent and illuminating answer to both the cynic and the sceptic. In sub-project (1), we develop an answer to the cynic according to which the kind of experience emotions are, and the kind of understanding they exemplify, is such that it makes no sense to think that these are just the personal rewards we reap from caring, rather than the positive orientation towards others. In sub-project (2), we explore an answer to the sceptic by making a case for the idea that the role of our emotions is not to determine what the right ethical decision is, but to anchor our emotional resources and energies within domains of interest, life projects and cherished values. If that is right, then we can explain what makes many cringe at the recommendations of the most vocal effective altruists (Singer, 2015; MacAskill, 2015, 2022), hopefully providing different ammunitions to the Williams-inspired frustrations elicited by effective altruism (Gray, 2015; Sreenivasan, 2015; Krishna, 2016; Adams, Crary, & Gruen, 2023), while at the same time responding to the very legitimate concerns it keeps on raising (McMahan, 2016; Greaves & Pummer, 2019).

## **2. Current state of your own research**

We hypothesise that looking into the nature of emotions and affective life more generally allows for a consistent and illuminating answer to both (1) the cynic and (2) the sceptic in matters of the importance of emotions for benevolence. What one might mean by "affective life" may vary enormously; in the context of the present project, it takes the form of three tenets that concern, respectively, (a) the nature of emotion, (b) the nature of evaluative concepts, (c) the nature of emotional sensitivities. These tenets have been elaborated over the last fifteen years in tandem with my colleague and co-author Prof. Fabrice Teroni. I now proceed to describe them.

(a) Any type of emotion seems to be characterized by the following features (Deonna & Scherer, 2010). First, it is intentional, meaning that it is directed at an object or state affairs: I am angry at someone or at the fact that I was ignored. Second, it has a formal object (Teroni, 2007). I can be angry at many things, but it is the offensive or unjust nature of the thing I am angry at that makes me experience anger and not another emotion. Third, different emotions feel in typically distinctive ways. Feeling anger is not the same as feeling amused or feeling fear. Fourth, emotions have action tendencies. We tend to react in certain ways when angry that are very different from the reactions we tend to have by other emotions. Finally, emotions are associated with evolutionary functions. Creatures capable of getting angry in adverse circumstances put them at an advantage. Not all theories of emotions in philosophy will insist on all these features and we find that different theories tend to stress a mere subset of them. Still, we may say that the majority agree that a maximalist theory of emotions will do justice to all these features, and, in particular, to the second feature, which embodies the idea that emotions have a deep connection to values (Deonna & Teroni, 2014). Anger relates us to offense, fear to threats, amusement to the

comical, shame to degradation, etc. The bulk of recent debate in the field has been concerned with how best to cash out the relation we have to values in emotion. One of the main models has it that emotions *represent* values in an experiential way. Anger, say, would be a phenomenological salient way of representing offense; unfitting anger would, e.g., be experiencing offense where there is none (Milona, 2016; Tappolet, 2016). Another model, one that our group has been pushing for the last decade, has it that emotions are not representations of values but rather *attitudes* or *stances* we take towards objects or state affairs (Deonna & Teroni, 2012, 2015, 2021). Anger is the attitude that we take towards objects or state of affairs that are offensive and it is as such that they qualify as evaluations of the world – not because they represent values. The sense in which emotions are different evaluative attitudes, and this is the first tenet of what we believe affective life consists in, resides in the fact that they are different felt action-tendencies. In anger, we experience *our own attention and body being engaged* in hostile fashion. Whereas the representational theory of anger says that these emotions are fitting when the situation is as they represent it, evaluatively speaking, the present theory has it that the felt hostile engagement is fitting when the situation facing the subject merits such an hostile engagement, that is, when it is offensive (Deonna & Teroni, 2021). In a slogan, the first tenet is that natural emotions are more or less fitting felt forms of attentional and bodily engagements. To say that they are “natural” is to say that they are pan-cultural psychological kinds, whose nature and concerns are not settled by cultural construction or evaluative reflection but, to a considerable extent, are given by human nature.

(b) The second tenet of our view has it that our understanding of evaluative concepts, or at least those that have an obvious link to affects, have an intimate relation to the emotions conceived on the attitudinal model (Deonna & Teroni, 2017). The concept of the offensive is rooted in anger, that is, is rooted in that which makes feeling engaged in hostile ways fitting. The concept of a wrongdoing is rooted in guilt and indignation, that is, in that which makes feeling engaged to repair the harm I caused fitting (guilt), or that which makes feeling engaged in having a harm recognized and redressed fitting (indignation). These are the examples we have in mind when explaining the importance of emotions and the way we *understand them* in our acquiring and mastering the evaluative concepts we apply to evaluate the world generally and ethically more specifically (Deonna & Teroni, 2020, 2021). Values and the related concepts are numerous and varied, as numerous and varied as the fundamental human concerns that dispose us to emotion and action. The circumstances we confront are rarely simple, and the *plurality* of values they exemplify are arguably difficult to compare, often making demands that compete or conflict, and in many cases, no obvious ordering of them in terms of importance is in the offing (Gill, 2012; D’Arms and Jacobson 2023, Ch. 8). Important for the present project is the distinction between thick affective *personal* and *impersonal* evaluative concepts. Arguably, threats, losses and achievements are ways of being bad or good *for* some individual or group, while beauty, fairness, happiness and suffering are ways of being good or bad *tout court* (Rønnow-Rasmussen, 2011). Both types of evaluative concepts, those that are directly related to the welfare of particular agents or groups of agents (fear, sadness, resentment, pride) and those that are not (admiration, indignation, sympathetic enjoyment, being moved, compassion) may claim to have roots in our understanding of various fitting emotional attention and engagement with the world.

(c) The third tenet has it that our emotions are the manifestations of our emotional *sensitivities* and that these sensitivities constitute the core of who we are, or, in other terms, of our identities. As suggested above, our emotional sensitivities are structured around a plural but finite set of basic human concerns: our sense of fear,

disgust, anger, admiration, envy, shame, etc., are structured around the (dis)value of certain aspects of our environment. Psychologists often refer to these values as “core relational themes” (e.g., Lazarus 1991) in order to emphasize their connexion to basic concerns. We may think of emotions as being at first responses to what de Sousa (1987) has aptly described as “primitive scenarios”, i.e., recurrent and basic situations that emotions have been set up to deal with. The fact that people react differently to different things in their environment is to a large extent due to their specific sensitivities, the dispositions that structure and organise their distinctive felt affective life. Crucial components of our sensitivities are *personality traits* and *sentiments*, both of which manifest themselves in consciousness through patterns of attention and experienced emotions (Deonna & Teroni, 2009, 2012; Watzl, 2017; Teroni 2023). The former is a multi-track disposition reflecting the weight that we, through normal acculturation or more idiosyncratic deliberation, put on certain specific values. The fair person’s concern for justice manifests itself in the way she reacts emotionally to the presence or absence of justice around her. She will feel indignation when injustice is present and guilt when she has inflicted injustice, etc. The latter are multi-track dispositions structured around the import objects, groups, institutions, etc. has for us. The love we feel for our child, country or charitable cause manifests itself in all sorts of emotional reactions depending on how the child, the country or the charitable cause fare. These emotional episodes we go through, the kind of *attention* they mobilise and the *engagements* they prevent or promote constitute the manner in which these sensitivities occupy consciousness. It is natural to think that the core of our identities as individuals engaged in our particular world is constituted by these emotions, together with the reflective attention we give to them in the process of determining whether the circumstances are such as to merit those emotional reactions as well the attempts at regulating them (Deonna, Rodogno & Teroni, 2011, pp. 88-90, Deonna & Teroni 2013).

Finally, let us emphasise some research accomplished over the years that run though all three tenets and is directly relevant for this project. First is the work on the nature of empathy (Deonna, 2001, 2007) that is obviously crucial background when researching affective benevolence. Second is the experience of developing focused and detailed research on specific emotions that is at the intersection of philosophy of mind and moral psychology. Note in particular the relevant work on shame (Deonna & Teroni, 2009b, 2011; Deonna, Rodogno, & Teroni, 2011) and on being moved (Cova & Deonna, 2014; Deonna, 2020), the latter being central in the exploration of the place of impersonal value as potential trigger of emotion. Third is the work on emotion and well-being (Deonna & Teroni, 2013, 2022b), in which we defend one version of so-called “hybrid conceptions”: the more we are given to engage emotionally with what is worth pursuing, the better our life is.

Although much more would need to be said to complete the picture, the affective life depicted by the three tenets taken together combines to form an ethical framework that in the following we shall call “*sentimental pluralism*”.

Lastly, it is essential to observe that the aims of this project will be pursued in an environment that combines the expertise and strengths of the Genevan Centre for Philanthropy (GCP), the Swiss Centre for Affective Sciences (CISA), as well as of *Thumos*, the Genevan group for research in emotions, values and norms, which I run in my capacity of Full Professor occupying, together with Fabrice Teroni, the chair for the philosophy of emotions in the philosophy department of the University of Geneva. I am lucky enough to be project leader in both the GCP and CISA, that is to say, in centres that are at the forefront of the research advancing our knowledge of the subject matter of the present project.



### 3. Detailed research plan

In sub-projects (1) and (2), we harness sentimental pluralism to answer to both the cynic and the sceptic. Sub-project (1) ties together tenets (a) about the nature of emotions and (b) about the nature of emotional evaluative concepts, to answer the cynic. The overall approach consists in focusing on the epistemic rather than the motivational credentials of emotions, and more specifically to understanding, an experiential kind of understanding, as being the epistemic good that we aim to uncover. The methodology pursued is one in which attention is paid to particular emotions. These include (i) guilt and (ii) compassion and what we shall call (iii) “sympathetic enjoyment”. The doctoral student hired will work on (ii) and (iii).

(i) Consider guilt: the idea is that an emotion like guilt is not just a bad feeling whose elimination can be achieved in any available way. It is an experience of a harm one has committed as needing to be repaired (Vazard & Deonna, 2019). If that is the case, then the cynic’s core idea that the help we lend to another is a selfish action aimed at easing our own pain becomes more difficult to sustain. Indeed, the very awareness of what the interpersonal configuration is like in an experience of guilt is such that, in the typical case, we do not want to describe this as a subject pursuing her own self-interest. The pain the subject feels in guilt is the pain felt at the harm she has inflicted and therefore a pain at a harm that can be undone only in a handful of ways, namely, by redressing or apologising. In short, the subject is emotionally responding to a reason to act in a guilt-adjusted manner. Our hypothesis is that we have here an explanation from the philosophy of emotion as of why it is right to say that the cynic’s description of what happens in guilt does not meet the phenomenal *cum* normative facts (see Elster, 2011). Developing the hypothesis to the full of course implies giving a convincing description of the *experience* of guilt as a response to wrongdoing, that is, a response to a reason to act. It will also require meeting the following challenges. First, while we may be right that experiencing guilt feels like aiding in response to a wrongdoing one has committed, the cynic will go insisting that the subject might still be deceiving herself in taking the situation as one in which what we are after is the well-being of someone else. Second, there is a quite common picture of guilt according to which what bothers us in guilt is not so much the harm we inflict on others through our wrongdoing, but rather the harm we inflict to ourselves by having turned ourselves into wrongdoers (Gangemi & Mancini, 2011). Perhaps guilt is primarily concerned with the displeasures of being immoral and the gratifications accompanying the restoration or the signalling of our moral standing (Bodner & Prelec, 2002; Holton, 2016). If that is true, then the cynic might be quite right in thinking that all we ever do is advance our own interests, while, more often than not, being under the illusion that we do something else (Simler & Hanson, 2017). In response, we anticipate that the correct reply is in both cases partly to concede to the cynic but refuse her more general conclusions. First, yes, we do sometimes self-deceive, but there is no easy route from this fact to the thought that we systematically self-deceive. Second, yes, guilt sometimes takes the form of signalling one’s virtue or of beating one’s breast about something where concern for the other seem not to feature in the experience at all. Yet, as close attention to the phenomenology of different types of preoccupations in guilt will reveal, including the phenomenology of dissonance when concern is about self-image (Aronson, 1992), it would be a mistake to think that these are the central cases (e.g. Wolf, 1982; Levy, 2023).

(ii) Consider now compassion. The idea here is to explore this emotion in connexion to, or perhaps in contrast with, the mere pleasure of benefiting others, irrespectively of the fact that they suffer, what we call “sympathetic

enjoyment.” Compassion has always been of great interest to moral philosophers (Aristotle, 350BCE/1926, *Rhet.* I385b13 ff.; Mandeville, 1714/1997; Hutcheson, 1725/1991; Hume, 1751/1998; Blum, 1980; Snow, 1991; Nussbaum, 1996; Crisp, 2008; Archer, 2018). There has been much less focus on it from philosophers of emotions, however, let alone from a perspective detached from the role of compassion in issues surrounding benevolence (but see e.g. Ben-Ze’ev, 2000; Tappolet 2000, 2023; Roberts, 2013; Kurth, 2019). The present investigation will focus first on the nature of compassion, an area in which much uncertainty and confusion has been brought lately. For example, does compassion have negative valence, as it was always thought to have, or does it have positive valence, as some recent research seems to claim (Klimecky & Singer, 2015; Bloom, 2016)? What is the formal object of compassion? Is it suffering? Is it undeserved suffering? If so, is that not too complex to be a basic natural emotion? The hypothesis that we shall explore in this part of the project is that compassion is *an experience of someone else as suffering*, which is itself a form of suffering. To gloss this account in terms of the attitudinal account presented above, this would mean that compassion is the experience of someone else’s condition as needing to be alleviated. That is the kind of *specific first-personal understanding* of the plight of others that compassion offers. Not only does this seemingly prevent the cynic’s reading to get off the ground, but also constitutes a basic and fundamental entry into others’ lives through their unhappiness. Now, this account opens a whole series of controversial issues. What about compassion felt towards people that do not seem to suffer? If compassion is hedonically negative, what is the difference between compassion and empathy felt towards suffering subjects (Nilsson, 2011)? If the formal object of compassion is simply suffering, does it mean that it is fitting to feel compassion towards people even when they deserve their suffering? A crucial feature that this project wants to bring to the fore is that suffering (like beauty, justice, excellence) is an impersonal disvalue with which we are equipped naturally to respond (see Gummerum, Takezawa, & Keller, 2009 for some supporting evidence; Nussbaum, 2015, ch. 6). Contrary to the Strawsonian wisdom in the area (Strawson, 1962), it develops alongside, and not as a form of sophistication of, our equally natural sensitivity to the personal versions of these values (my beauty, my being treated unfairly, my excellence).

These questions will be taken up along another concerning the emotion we feel when we experience or anticipate someone else’s happiness, independently of any unhappiness in which they might or might not be. Along those lines, we often say that we are “happy for” someone, in a form of sympathetic enjoyment. These positive feelings we experience in relation to the felt happiness of others resemble a great deal the emotion discussed under the label “compassion” (sic!) by the recent trend that tries to put it as the emotion we should be treasuring, as opposed to “empathy,” even to the point of portraying it as a hedonically positive emotion (Bloom, 2016; Klimecky & Singer, 2015). The famous “warm glow” feelings posited by James Andreoni (1990) in the context of philanthropic giving seem to be another instance of this phenomenon. And, in a completely different context, the “gratulation” experience described by Kristjánsson, which is felt at the deserved good fortune of others, seem to be in the same ballpark (Kristjánsson, 2015). In the case of the warm glow, this feeling seems to be linked to the good fortune *we* as benefactors bring to the beneficiary, while in the gratulation case it seems to be the pleasure felt at the good fortune the beneficiary has managed to bring on himself (Bianchi, Cova, & Tieffenbach 2023). One of the key questions in the area is whether there is a natural emotion (D’Arms & Jacobson, 2023) around the happiness felt at the happiness of someone else, which is (1) different from any other mundane pleasures we might feel at any old good occurring or (2) pleasures not necessarily associated with the fact that suffering has been

eliminated and which (3) is impersonal in nature, in the sense that the other person's identity in relation to me, or the fact that I am the benefactor or not, does not matter. If there is a natural emotion in this area, we may surmise that this emotion provides a fundamental understanding of others as pursuers of their own happiness. If sympathetic enjoyment and compassion are fitting responses to happiness and unhappiness respectively and are the expression of fundamental human concerns, then perhaps they play a crucial role in extending our mental life beyond the sphere of our narrow attachments. This would have important implications as well for morality in general – more on this below.

(2) This sub-project is focused on responding to the sceptical route to the sidelining of emotions in the context of benevolence, especially as it plays out in the field of the ethics of philanthropy. For the reasons adduced above, effective altruists have been pushing the sceptical route vis-à-vis emotions (MacAskill, 2015; Singer, 2015). Emotions, not being fit for delivering verdicts that would cater for the good of everyone, should be forsaken. Opponents, in particular those inspired by Bernard Williams' brand of pluralism (Williams, 1985), have reacted with dismay (e.g. Krishna, 2016; Hoel, 2022). Affective altruists, they insist, through their advocacy of "impersonal, ruthless decision-making, heart firmly reined in by the head" (Srinivasan, 2015) are simply asking people to repudiate their identities. This, we are told, is not only ludicrous psychologically speaking, it is also ethically unacceptable. In this part of the project, we try to rise above the battlefield and enrich the sentimental pluralist approach to the ethical so as to respond to the sceptic. We do that while trying to show that we can, from within the sentimental pluralist approach, do justice to the effective altruist's most pressing concerns.

The sensitivities described above, chief among them our personality traits and sentiments, are the result of a process that is contingent on our immediate environments, namely, the people (caregivers, friends, peers), groups (family, clan, community, institutions) and values (variety, level of importance) in which we happen to grow up and around which live (de Sousa 1987, D'Arms & Jacobson 2010). When they crystallise, these emotional dispositions track the fate of the people, institutions and values we are attached to; the emotional responses they raise are the manifestations of these attachments. As far as we can and especially for those of us who have the luxury to do so, we try to stay alert or attentive to these people and those things we care about (Murdoch, 1970/2002; Ouellette-Dubé, 2022). As argued above, inasmuch as our emotional dispositions and their manifestations are the reflexion of what we love and hate, they constitute the core of our identities. In expressing our benevolent inclinations, then, it is no surprise that these emotions will cater first and foremost for the people, institutions and values that are in line with these identities. Overall, we think that this is good: we value the fact that people welcome, guard, fear and hope for what is at the centre of their lives. We see instrumental value in these identities of course, but also and especially their intrinsic value, both for oneself or for others (e.g. Bluestein, 1991; Wolf, 2010; Wonderly, 2016). If that is the case, then it is at this juncture that the fundamental tension around which one of the central debates about benevolence is structured.

On the one hand, the very nature of our identities is such that *partiality* is the name of the game (Scheffler, 2010b; Jollimore, 2011; Keller 2013). We have attachments to people, projects and values that we rightly believe make for the meaning of our lives. That they should be prioritised is exactly how it should be. Privileging the people, projects and values we love is to be fostered not because it is an inescapable feature of human nature – rule consequentialism cannot come to the rescue here – but because it is what makes our lives intrinsically valuable

(Stocker, 1996; Kauppinen, 2011; Deonna & Teroni, 2013, 2022). Forsaking one's identity in favour of treating all people and their projects alike, says the sentimental pluralist, is alienating and tantamount to rejecting who we are altogether. This, we hypothesise, is true of our everyday experience in the social world, and it is still true when we come to engage in more organised philanthropic endeavours.

On the other hand, as previously diagnosed, if this defence of partiality is good for ourselves, then it must be so for others too, including those that are not in our sphere of attachments. We are here facing one expression of the so-called "distributive objection" (Scheffler, 2001). The latter individual, insists the effective altruist, must be allowed to flourish within the space of their attachments and projects too. If this is recognised, as it should, provision for the possible expression and flourishing of other people's identities should be catered for. Thus, sentimentalists should cease finding excuses in human nature to ignore the suffering of those they cannot see (McMahan, 2016; Greaves & Pummer, 2019).

The present project explores different ways to rise above the present stalemate and hopes to delineate the contours of *permissible* and *required partiality* (Lange, 2022) with an eye on philanthropic giving. It aims to do that in a framework that provides more than usual in matters of what grounds partiality, that is, our identities understood as emotional sensitivities as well as answering questions pertaining to the fate of people whose existence seem to fall under the attachments of no one person or community.

(a) A first task is to integrate the framework offered by the affective life of the three tenets with the existing developments on attachments to *projects* (Wolf, 2015), *relationships* (Kolodny, 2010) and *individuals* (Keller, 2013) as they surface in the literature on the sources and justification of partiality. An important worry in the area concerns the fact that projects and relationships may be thought to be valuable only *derivatively* on the value of the projects and people we are attached to (Keller, 2013). If that is the case, then we may fear that the sentimental pluralist has nothing of intrinsic value to point towards in order to justify the partiality she wants justification for. One of the arguments that will be developed in this context is that our emotional attachments, though not themselves *sources* of value – they are what some have aptly called "value *enablers*" (Jollymore, 2011; Rabinowicz & Rønnow-Rasmussen, 2021; Teroni, 2023) – do in fact yield intrinsic value themselves (Betzler, 2014; Olson, 2014) when they constitute the kind of *fitting* orientations to their objects that we have described above (see tenets (a) and (c)).

If that is granted, then the intrinsically valuable deployment of attention and engagement that these emotional interactions involve will allow for the kind of flourishing we want for our lives. Yet, it will also necessarily tend to set aside or "silence" (McDowell, 1998) other projects, relationships and values however intrinsically valuable they may be. If this is accepted, as it should, do we have the resources to remedy the negative fallout within the sentimental pluralist framework? More on this below.

(b) In the meantime, note that philanthropic organisations and charities raise particularly interesting issues in relation to their status within our attachments. At one end of the spectrum, philanthropic organisations are seen as surrogates for what elected governments cannot achieve due to institutional regulations, lack of resources or malfunctioning, etc. If so, then our attachment to philanthropic organisations would be close to the kind of

attachment we have (or do not have) for the state (to which, as things go, we have to pay taxes). At the other end of the spectrum, they are viewed as specific projects and/or values people pursue as part of their identities (Boesch, 2018; Swanton, 2018; Crary, 2023). This way of being engaged for the collective good would perhaps complement the state but not substitute for it. One hypothesis we shall explore here is that the effective altruism movement, on account of the fact that it asks us to treat everyone's project alike, is at best turning everybody into little governments that must have the interests of all as their only objective (or, at worse, calling on us to volunteer paying more taxes). Taken at its limits, we shall argue that this is a mistake.

Armed with the concept of identity developed above, we hypothesise, against recent pessimism (Reich, 2018), that philanthropic pursuits should be conceived as the sphere in which the plurality of values embodied in heterogeneous liberal societies can express themselves at the collective level (Tyler, 2019). Because philanthropic organisations embody values of a certain type, our commitments to them embody the expression of our values at the collective level and, in favourable circumstances, allow for projects and values that is not the prerogative of the state to find adequate expression. If that were the case, then the goodness of philanthropy would not lie in the efficiency with which it might cater in equal fashion for the most needy, but rather in the way it provides opportunities for people to project their attachments while contributing at the same time to their own flourishing and to the collective good.

Many challenges will have to be met in the process of developing this argument. The avowed ethical pluralism in which the present project is embedded will present considerable challenges (Gaut, 2002; Zagzebski 2003; Kauppinen 2013, 2015). Chief among them is again the worry that the sentimental pluralist will not be in a position to take into account the fact that people should be given the opportunity to express and pursue their own identity even when they are not in the spheres of anyone's attachments. As a veteran friend of partialist morality would remind us, "no moral outlook [is] acceptable if it fails to account for the norms governing our treatment of distant strangers" (Scheffler, 2010b, p. 128).

To rise to this challenge, the following more comprehensive picture of our identities will have to be developed. A *mature* identity, in the spirit of the pluralism stemming out from the richness of our fundamental concerns, will be constituted by dispositions that heed the existence of a rich palette of the most important evaluative properties: social and non-social, moral and non-moral, personal and *impersonal*, etc. – beauty, excellence, friendship, knowledge, safety, but also justice, suffering and happiness. Through sensitivity to the last three, just to mention them, people should detect plenty of reasons to act in ways that mimic the effective altruists' desired state of affairs in a way that is grounded *not* in a universal point of view, *not* in wanting to do the most good, but in their fitting dispositions to respond to what is (dis)valuable in this or that. Fitting indignation at certain injustices makes *pro tanto* demands that we help in faraway places people that we do not know because our responsibility is or was engaged (Pogge, 2011). In other cases, our compassion for the suffering of these same people will make the same *pro tanto* demands irrespectively of considerations of justice (Lomasky, 1995). In still another cases, the anticipated sympathetic enjoyment of the happiness of total strangers (see sub-project (1), (iii)) will be enough to make comparable demands, and, hopefully, anticipatory benevolent gestures.

Talk of mature identities is potentially both naïve and dangerous. It is naïve because it may seem preposterous to think that we might even come close to decide what these would be like. It is dangerous because it might conjure up thoughts of paternalistic approaches to education or, worse, social engineering experiments. Still, it makes sense to investigate the idea that, although there are many different ways to realise a mature identity, none will be ideally mature that leave important impersonal social values fully unprotected or never promoted. One hypothesis that we will explore in this context is the idea that we can train our emotional skills and energies to be more responsive to these values in particular, a goal that will fit well within research in education studies that has finally started to receive the attention it deserves (see Kristjánsson, 2018, ch. 9 for a review). If we believe evidence according to which younger children's benevolent inclinations does not especially favour in-groups (Gummerum, Takezawa, & Keller, 2009), then it might involve preventing losing existing dispositions rather than inculcating new ones. The promise contained in this suggestion is that it would be a way of harnessing the kind of understanding that emotions provide at the service of more universal concerns, a moral education that keeps the growing subject watchful of the more universal demands of morality without these demands feeling anything other than emancipatory (e.g. Nohra, 2006; Ouellette-Dubé, forthcoming). In the spirit of what we want to defend in this part of the project, the goal is never to turn people into "benevolent bureaucrats distributing [...] benefits in a primarily administrative" role (Nagel, 1972). Rather, the goal is to make them understand that those to whom we are not attached should also be given a chance to have life. This may be done exactly through the emotions that are the expression of the ordinary palette of human fundamental concerns for what is good for oneself *and* good tout court.