Falling in Love and Breaking Up: Attribution Bias and the Perception of Responsibility

Michelle CIURRIA
University of Missouri - St. Louis
mciurria.academia@gmail.com

Abstract: I argue that our tendency to self-attribute or 'take responsibility for' positive rather than negative events explains why we tend to experience romantic love as responsibility-preserving and painful break-ups (in many cases) as responsibility-undermining. This seems to be the case even though both experiences share similar phenomenological qualities, particularly lack of control and loss of psychological continuity. I contend that our asymmetrical perceptions of responsibility in positive versus negative cases can be explained as an effect of a properly functioning attribution-self-representation system, which generates differential attributions in the two cases. I suggest that the subjective perception of responsibility – supported by our attribution biases – might be relevant to theorizing about how we attribute responsibility to people. This implies that we should withhold blame from people with a pessimistic attribution style, and withhold praise from people with a narcissistic attribution style, as a way of bringing them closer to the average range. This is because the average degree of optimistic bias has adaptive value: it makes us happier and more functional. So, the subjective perception of responsibility and the objective reality of responsibility are intimately related. Finally, I show that gender influences self-attribution style, which has implications for how we should hold men and women responsible.

Keywords: responsibility, romantic love, control, narrative identity, attribution theory, agency enhancement theory.

1. Introduction

Most people tend to attribute positive events to themselves and negative events to external causes. That is, they tend to perceive themselves as being more responsible for positive than for negative events. This is exemplified in the experience of romantic love. Romantic love, according

to Alain Badiou (2012) and Alain de Botton (2016), has an essential element of chanciness: we *fall* in love, we don't choose to love someone. Similarly, when we break up with someone, it can (depending on the circumstances) feel like an accident, a tragedy – something over which we have no control. Yet falling in love does not seem to threaten our feeling of being responsible – of being in control, of being a complete person – in the same way that breaking up can. This points to an asymmetry in how we perceive our responsibility: we feel more responsible for positive events than negative events (for falling in love than breaking up under infelicitous circumstances), even if both experiences have similar 'chancy' (or deterministic, as the case may be) qualitative features.

Why do some people take responsibility for positive events more than others? I will suggest that this is a function of a person's attribution-self-representation system, which generates a nexus of causal attributions and self-reflexive beliefs. Some people have a pessimistic attribution style and corresponding negative self-representation beliefs (especially people with depression); others have a self-aggrandizing attribution style and corresponding egoistic self-representation beliefs (especially narcissists); and neurotypical people (neither depressed nor narcissistic) make moderately optimistic attributions and moderately positive self-representation beliefs. Research shows that narcissistic individuals self-attribute successful outcomes more than less narcissistic individuals do (e.g., Stucke 2003; Brown 2013), people with depression self-attribute negative outcomes more than others (Rubenstein et al. 2016), while neurotypical people lie somewhere in the middle. The ordinary person's moderate degree of self-serving bias has adaptive value: it tends to facilitate happiness and social functioning. This is why ordinary people tend to live 'ordinary' lives, neither fraught with depression nor plagued by self-defeating narcissism.

Does a person's subjective perception of responsibility, produced by her attribution-self-representation system, have any implications for how we *ought* to think about responsibility in theoretical terms, or for how we *should* attribute responsibility in practice? It might, if responsibility attributions are supposed to influence people's behavior. On a forward-looking, 'agency enhancement account' (see Vargas 2013, McGeer 2013), it makes sense to think that our expressions of praise and blame should respond to a person's attribution style and self-representation beliefs, 'nudging' those beliefs toward the average if they're excessive in either direction.

In section 2 below, I will briefly present an account of responsibility, to be used further on

in the analysis. In section 3, I talk about how romantic relationships exemplify our tendency to take responsibility for positive events more than we do for negative events. In section 4, I suggest that this tendency can be explained by reference to an attribution-self-representation model of positive psychology. In section 5, I argue that a person's attribution-self-representation style has implications for how we should theorize about responsibility as an interpersonal practice. And in section 6, I highlight the intersections between gender and attribution style, and the implications of these intersections for agency-enhancing models of responsibility.

2. Responsibility

Philosophers use the word 'responsibility' in many different senses. Some of these senses imply that what we are responsible for depends on what we experience as authentic, under our control, and determined by our choices, as opposed to alien, uncontrollable, and imposed from without (whether by chance, coercion, or duress). David Shoemaker claims that on the 'attributability view' of responsibility, what we are responsible for are the traits, choices, and states of affairs that we would label as "authentic" as opposed to "alien" (2014: 120), while Fischer claims that on the 'control theory,' what we are responsible for are the traits, choices, and states of affair that are under our control or "up to us" (2012; cf. Scanlon 1998). Notably, whether a trait, choice, or state of affairs *appears to be* attributable to us or under our control depends on our attribution-self-representation style, which, in neurotypical people, represents positive states of affairs as 'ours' and 'chosen by us,' and negative states of affairs as alien and coerced. In narcissistic and people with depression, the attribution-self-representation system works very differently. Thus, our perception of our responsibility will depend heavily on our attribution style. This does not mean that whether we are responsible depends on our subjective psychology - but our *perception* of our own responsibility does.

Other theorists define responsibility as a system of reactive attitudes, such as blame and praise, which are justified by their propensity to enhance the agency of the recipient. Manuel Vargas' 'agency enhancement theory' is an example (2013). On this approach, whether one should hold someone responsible for a choice does not depend on whether one regards that choice as her own or under her control; it depends on how holding her responsible would affect her agency.

In what follows, I will argue that falling in love and breaking up often share similar phenomenal features incompatible with responsibility – such as chanciness, uncontrollability, and alienation – yet due to our attributional biases we don't take responsibility for both kinds of events equally. Furthermore, we *should not* take equal responsibility for both, because ordinary attributional biases are adaptive and healthy.

3. Love, break-ups, and asymmetrical perceptions of responsibility

The features that falling in love and breaking-up share – perceived loss of control and loss of psychological continuity – are in tension with dominant paradigms of responsibility, which link responsibility to control (e.g., Fischer 2012) and the possession of a persistent and coherent ('deep') self (e.g., Frankfurt 1871, Sher 2010). When we fall in and out of love, then, we should experience a disorienting loss of responsible agency. But only in the negative case – the break-up – do these features challenge our *feeling* of being a deep, continuous ego. This asymmetry, I submit, can be explained by an optimistic attribution bias found in neurotypical psychology: we tend to self-attribute positive events (like romantic love), and to externally-attribute negative events (like break-ups). We can see this asymmetry exemplified in some popular philosophical accounts of romantic love – specifically, Alain Badiou's (with Truong, 2012),[1] and Alain de Botton's (2016), to which I turn to now.

According to Badiou, one of the defining features of romantic love is that it emerges spontaneously out of a chance encounter. It cannot be "conceived… as an exchange of mutual favours, or … calculated way in advance as a profitable investment… love really is a unique trust placed in chance" (17). When we fall in love, it happens abruptly and unexpectedly, often as a result of a seemingly meaningless encounter; but this chance event quickly changes into something eminently meaningful – something that seems almost inevitable:

> *The chance nature of the encounter morphs into the assumption of a beginning. And often what starts there lasts so long, is so charged with novelty and experience of the world that in retrospect it doesn't seem at all random and contingent, as it appeared initially, but almost a necessity. That is how chance is curbed: the absolute contingency of the encounter with someone I didn't know*

---

[1] All subsequent references to Badiou are to 'In Praise of Love,' Badiou & Truong 2012.

finally takes on the appearance of destiny (Badiou: 42-43).

This may reflect our tendency to 'take responsibility' for the encounter; what initially seemed accidental and destabilizing in retrospect appears natural, voluntary, self-actualizing. Another salient feature of romantic love is that the self is transformed: the lover no longer sees the world from the perspective of an atomistic agent, but from the perspective of a 'duality'. "Love isn't simply about two people meeting and their inward-looking relationship: it is a construction, a life that is being made, no longer from the perspective of One but from the perspective of Two" (Badiou: 29). This marks a fundamental change of self-conception: you no longer live for yourself only, but for the intertwined interests of a couple. Your interests, values, and commitments are, not exactly *subsumed* into, but *mediated* through, those of your partner. This is a transformative experience.

These features of the phenomenology of falling in love are in tension with dominant paradigms of responsibility, which focus on *control* and *continuous narrative identity*. Chanciness conflicts with control; and radical self-transformation conflicts with self-continuity – with the persistence of a stable ego through time. If we were kidnapped and brainwashed, we would feel that our responsibility had been undermined by hostile external forces, as Robert Mele describes futuristic 'manipulation cases' (Mele 1995). Not so when we fall in love; we feel no less responsible than before. Yet oddly, the salient features of these two experiences – loss of control and self-transformation – are the same. Only in the negative case do they harm our self-conception.

Here's a possible objection to this conception of love. Maybe you think that you could choose a romantic partner on the basis of reasons, weighing the pros and cons of the person's various qualities, as you might order a meal from a restaurant menu. This would cancel out the disorientation and bewilderment that accompanies falling in love 'by accident.' But, according to Badiou, such a scenario (choosing on the basis of a rational decision procedure) is anathema to romantic love: it eliminates the risk that makes love meaningful and transformative; it reduces 'falling in love' to a mundane market transaction, something akin to taking out an "insurance policy" (9). Buying an insurance is, of course, an eminently responsible thing to do; but when it comes to romantic love, it defeats the purpose. *Love can't be responsible.* By its very nature, it is an adventure and a risk. Now, we might rationally decide whether or not to marry someone, and whether or not to have children with someone, but that's not how we fall in love. We don't choose,

we *fall*. It seems to be built into the modern western conception of romantic love (if not all conceptions) that falling in love is not exactly irrational (i.e., against reason), but *a*-rational (i.e., not decided on the basis of reasons). Love can coincide with reasons but cannot stem from them. Ordinary people tend to associate romantic love with excitement and sexual arousal. For a more 'rational' emotion one could look at companionate love instead (Brogaard 2015).

Now consider the flip side: breaking up with someone. Badiou doesn't write much about break-ups. He does say that "everybody knows that deciding to break off such love, particularly unilaterally, is always a disaster, whatever the excellent reasons put forward to support such a move" (46). And I think this is right inasmuch as breaking up is often the opposite of falling in love: a transformative disaster, something that harms your self-conception. Some break-ups are rational – for example, if your partner is abusive. In that case, there are very sensible reasons for breaking up. But other times, break-ups have similar experiential qualities to falling in love: they seem unreasonable and random, precipitated by contingencies or accidents beyond our control – an unanticipated series of unfortunate events. Or they can be brought about by a betrayal that calls into question whether the relationship was ever 'real' in the first place. In either case, the break-up can have the quality of what existentialists call 'the absurd': it resists sensible explanation, destroys meaning, shatters the very system of values that were constructed and solidified in relationship with the other. Similarly, a break-up can challenge one's sense of self as a stable ego continuing through time, a part of something bigger than oneself. That is, a break-up can undermine one's feeling of being in control and being a coherent psychological entity.

Other times, break-ups can happen because of psychological factors (seemingly) beyond one's control, such as an attachment disorder acquired in childhood. De Botton argues that we always marry the 'wrong' person because "we have all emerged from childhood with a bewildering array of disturbances that come into play when we try to get close to others. We can only ever seem normal to those who don't know us very well" (2016). When these developmental issues interfere with our adult relationships, they can feel like irresistible compulsions overpowering our will. Nonetheless, we can, says de Botton, bond with people in spite of our various "bewildering disturbances," provided that we are able to "tolerate differences with generosity," and we manage to find someone who can do the same. In some cases, however, personal differences can feel utterly insoluble, and this can give a break-up the appearance of an inevitability, though an utterly

unpredictable one. Desert-based theories of responsibility have a 'foreseeability' constraint, which rules out the possibility of being responsible for an unforeseeable outcome (see Fischer 2012). If so, then the impossibility of knowing another person prior to falling in love, and the impossibility of detecting subtle incompatibilities between the two of you, can defeat control, thereby defeating compatibility on the control model.

Falling in love and breaking up, then, can have similar experiential qualities: chanciness, lack of control, lack of self-continuity, self-transformation, unpredictability. But only in the negative case – the break-up scenario – do these qualities pose a threat to the phenomenal self. Only in the negative scenario do we feel helpless, impotent, diminished – qualities that are anathema to the perception of responsibility, the experience of being able to exercise control over one's life and the capacity to continue to exist as a whole person. While falling in love is like a rebirth or reconfiguration of the self, breaking up can be like a death. This is not just because the experience is unpleasant: root canals are unpleasant, but they do not threaten our sense of self. The abruptness, the loss of control, the loss of authenticity, are some of the properties that make the end of a meaningful relationship a personal tragedy. The identity that we inhabited with the other, that was *co-constituted* by the other, no longer exists. We are thrust into a world that does not make sense from the perspective of either a singularity – the identity we've abandoned - or a duality – the identity we've lost. We are forced, whether we like it or not, to adapt to being a different person, to finding a new way of being in the world and a new way of relating to others.

Why is there this asymmetry between our perception of responsibility for positive events versus negative events, for romantic attachments versus break-ups? Why do we experience love as enriching, and break-ups (at least, certain kinds of break-ups) as destabilizing, bewildering, defeating, even though the two experiences have similar responsibility-relevant features? I turn to this question in the next section.

4. Attribution biases, self-representation beliefs, and perceptions of responsibility

One way of explaining the relevant asymmetry is by appealing to attribution theory. Different people have different subjective perceptions of responsibility, and these seem to be tied to our attribution styles. Most people self-attribute, or 'take responsibility for,' positive events

more than negative events. There is research on this phenomenon (Mezulius et al. 2004),[2] but it can also be documented by a careful observer. Indeed, almost 4 centuries ago, Shakespeare remarked on this human disposition in 'King Lear':

> *This is the excellent frippery of the world, that, when we are sick in fortune – often the surfeit of our own behaviour – we make guilty of our disasters the sun, the moon and the stars; as if we were villains by necessity, fools by heavenly compunction.*

In other words, people tend to blame the world for bad things while taking credit for good things. This *self-serving bias* affects ordinary people's causal judgments. But it's not universal. People with depression exhibit the opposite bias: they tend to self-attribute negative more than they do positive events. For example, if a person with depression gets a bad referee report, she is more likely to think, "it's because I'm stupid," rather than, "it was a fluke" or "the referee didn't get the point." Anyone can lapse into a pessimistic attribution style in the face of adverse events that seem out of their control (Abramson et al. 1978). But this mindset is usually temporary. A distinctive feature of depressive psychology is that people with depression tend to assume their negative traits are internal, global, and stable: they are about the person, they affect all of the person's behaviors (not just specific behaviors), and they seem immutable (ibid). This makes people with depression susceptible to prolonged dysphoria, not just the circumstantial pessimism experienced by most people. Depression therefore tends to be particularly trenchant – just as happiness in ordinary people is fairly resilient.[3]

Attribution style is only part of this story, however. One's mindset – whether depressed or optimistic – is also influenced by factors such as self-esteem, memory bias, and self-standards: features of our 'self-representation schema'. A pessimistic attribution style actually predicts low self-esteem better than it predicts depressive symptoms, though the latter correlation is still significant (Tennen & Herzenberger 1987). This means that self-esteem plays a mediating role in depression. Also, people with depression tend to remember negative information about themselves more than they remember positive information (Greenier et al. 1999) – they have a pessimistic memory bias. And they tend to show major discrepancies between their actual self and their 'ideal self,' the person they think they ought to be. Thus, they hold themselves to inordinately

---

[2] Mezulius et al. find through meta-analysis that most people exhibit some degree of self-serving bias.
[3] I am not saying that it's necessarily *better* to be a neurotypical person, but I will address the question of whether being non-depressed is adaptive on page 15.

perfectionistic standards, thinking things like, "I should be able to please everybody." These self-representation beliefs interact with their attributions in "an attribution-self-representation cycle, [such that] changes in either inevitably lead to changes in the other" (Bentall 2011: 5294-5). People with depression thus tend to have a combination of negative self-representations and pessimistic self-attributions, which reinforce one another, whereas neurotypical people tend to have more optimistic self-attributions and more optimistic self-representation beliefs, which hang together in a fairly stable equilibrium. Although a person's attribution-self-representation style can "change detectably over time" (Bentall 2011: 5303), it predicts overt behavior while it's present.

Attribution-self-representation theory can help to explain why the ordinary person's experience of responsible agency is strengthened by a new romance, but challenged by a painful break-up, even when the two experiences have similar phenomenological features, antagonistic to theoretical paradigms of responsibility. The explanation is that we are more inclined to 'take responsibility for' romantic relationships compared to break-ups (and their immediate aftermath), insofar as the former are seen as positive and the latter are seen as negative. Thus, even if both types of event seem 'chancy' (or determined, depending on the case), and even if both types of event disrupt narrative continuity, we might have an inbuilt inclination to accept the results of a romantic encounter as 'our own' and as voluntarily chosen, while rejecting an unforeseen break-up as alien and involuntary. That is, after entering a romantic relationship, a neurotypical person is likely to appraise it as something over which she is responsible, whereas in the aftermath of a painful break-up, she is more likely to attribute the event to external factors. This asymmetry may be caused by a broader disposition to attribute positive events to ourselves and negative events to externalities. So, even if the originating cause of a romantic relationship – a chance encounter – was completely out of our control, we might come to re-appraise the encounter and its downstream effects – the loving relationship – as attributable to our 'self' and our 'autonomous choice.' Not so with an unwanted break-up: we are likely to ascribe the painful end of a relationship and its after-shock to externalities, not our own choices and character traits. This outcome is predicted by attachment-self-attribution theory. Thus, this theory can be applied to romantic love and break-ups to explain the asymmetry in our perceptions of responsibility in each case (ceteris paribus). Both have the same objective features of non-responsibility (uncontrollability, self-alienation, chanciness), but they are experienced differently depending on whether the event is seen as positive or negative.

In contrast, people with depression might feel more responsible for a break-up and less responsible for a new romance. The phenomenological experience of romantic love of a depressed person, that is, might be different from the average person's. While the beginning of the relationship might feel like a fluke or a miracle – something external – the end of the relationship might feel like the person's own fault. If so, then the person with depression might have difficulty assimilating the new romantic relationship into her self-representation schema, and achieving the 'two-and-not-one' mentality described by Badiou. (She might have trouble taking responsibility for the relationship *qua* positive state of affairs). The person with depression might also have trouble recovering from a break-up because of a failure to externally attribute the negative event, thereby taking responsibility for the negative outcome.

Alternatively, she might generate external attributions for *both* events, in which case she's likely to have a generally deflated ego. That is, she might not be able to take responsibility for either the relationship or its demise – or for much of anything – resulting in self-alienation. The idea of a global responsibility deficit in persons with depression is, in fact, consistent with Freud's characterization of depression as a kind of ego depletion (1984), and with more recent associations between depression and symptoms of withdrawal, apathy, and avolition (Adams 2001).

There is a third possibility: someone might take *too much* credit for positive events and not enough for negative events. This person would perceive the beginning of a relationship as being her own doing, and the failure of the relationship as being the other person's fault. This type of person is a *narcissist*, in the sense that her attribution style reflects a vain and egoistic sense of her own self-importance. It's debatable whether the narcissist is capable of experiencing love, since narcissism is likely to impair one's capacity to invest in another person for that person's own sake, thereby achieving the duality and mutual investment central to romantic love. The narcissist, then, might be existentially alone, incapable of attaching to another person.

The tendency to take proportionally more credit – but not excessive credit – for positive events compared to negative events is common. And this is arguably why most people experience romantic relationships as responsibility-preserving (or responsibility-enhancing), even if romantic encounters – the causal origin of the romantic relationship – defeat control and narrative continuity; and this tendency also explains why we tend to experience painful break-ups as responsibility-undermining, even though - in fact, precisely *because* - they have these

responsibility-defeating features. That is, our attribution biases explain our asymmetric perceptions of responsibility for relationships versus break-ups.

Interestingly, on some philosophical accounts of responsibility, if we can 'trace' an event back to a non-voluntary causal origin, we are not responsible for the outcome (e.g. Fischer 2012; Vargas 2005). This doesn't seem to hold for romantic love, however, because even if we can't take responsibility for the 'chancy' causal origin of love, we tend to take responsibility for its upshot – the romantic relationship. This seems to be because, when it comes to evaluating positive states of affairs such as romantic relationships, we don't see external causation or lack of control as responsibility-defeating. However, when evaluating negative states of affairs like break-ups, we do: we find external causation to be troubling *only in the negative case*. This perception, of course, is not a question of objective metaphysics – about whether agents have contra-causal powers or some other special metaphysical status (see Kane 1985). It is a question of *subjective perspective*. But subjective perspective might be relevant to 'holding responsible' in an objective sense, if praise and blame are supposed to promote functional behavior. I will discuss this in more detail in the next section.

Now, whether the average attribution style is 'right' is a fraught question. The fact that it is a bias reveals that it is epistemically flawed: it does not track objective reality. It reflects a prejudicial favoritism for oneself as the cause of positive occurrences but not negative ones. Arguably, the perspective of the person with depression is the more objective one: maybe it is right to think that we are not responsible for romantic relationships, since they stem from a chance encounter. Maybe we are actually not responsible for anything – which is a position held by some incompatibilists ('hard determinists'). Interestingly, people with depression are known to exhibit 'depressive realism': the tendency to make (depressingly) accurate self-appraisals (see Bantall 2011: 16643). This might make their perspective more accurate: they are better at objectively assessing themselves relative to others. But these questions about objectivity, metaphysics, and epistemic authority are, in an important sense, beside the point. Whether or not the ordinary person's self-serving bias is 'objective' (or metaphysically sound, or what have you), there's no doubt that it is *adaptive*. That is, it promotes subjective happiness and social functioning. There is good reason, then, to promote this mindset. And this means that there is good reason to promote attribution biases that support this mindset.

In relation to this, it is also notable that people with depression and narcissists, according to what was said above, might not be able to experience romantic love (at least not to the fullest extent). People with depression might have trouble assimilating the relationship into their self-conception (due to pessimistic biases that prevent them from 'taking responsibility' for positive events). Narcissists, on the other hand, might be practical solipsists – they might not be able to bond with others and incorporate others into their ego. If so, then both biases are antagonistic to love. Thus, if we value romantic love and wish to promote it, we have a pragmatic reason to promote the average propensity for optimistic bias. So our responsibility attributions – praise and blame – should aim to promote the average (functional) level of self-serving bias. I will expand on this thought in the next section.

5. 'Real' responsibility?

What, if anything, does all of this have to do with whether someone *is* responsible? I've been talking for some time now about love and the perception of responsibility. But perceptions aren't real: they are mental representations. Responsibility is a real thing, right? Either we are responsible for something or we are not. We might *feel* more responsible for positive than for negative events (unless we are depressed), but that has nothing to do with whether we are *really* responsible for anything – so the argument goes. There's some sense to this reasoning. If someone cuts me off in traffic, and I think that he's to blame, but the driver insists it wasn't his fault, surely his subjective perspective can't decide the matter. Nor can mine. This is trivially true: objective questions of responsibility can't be decided by someone's subjective opinion. But this is separate from the question of whether subjective features of agents – features like attributional biases and self-representation beliefs – can be *relevant* to whether someone is an apt target of praise or blame, objectively speaking. Most theorists, in fact, assume without hesitation that the psychological properties of agents – their capacities, or their quality of will – are the defining features of responsibility. People are responsible just in case they have these properties. It's not especially controversial, then, to suppose that a person's psychological properties can affect her aptness for praise and blame. But it might be controversial to suggest that a person's *attributional biases* are relevant, since it's not clear that these states are related to a person's moral capacities in the right way. Yet, at least some philosophical theories – the ones that I take to be the most compelling –

seem to accommodate the idea that attribution biases *should* inform our praising and blaming policies and attitudes toward responsibility.

In particular, I think this idea fits nicely with Vargas' 'agency enhancement theory' (2013), a forward-looking account on which responsibility attributions (praise and blame) are 'fitting' just in case they're likely to enhance the target's moral agency. Roughly, we are meant to blame or praise people in such a way as to influence them to be better people. On this view, it's compelling to think that subjective attribution biases should be relevant to responsibility as an interpersonal practice, since these subjective biases and beliefs affect whether certain kinds of moral address are likely to be motivationally resonant, and thus whether they're fitting. This is an *objective* matter, since it's objectively true that some moral reactions will be more effective than others.

Here's how attribution theory interacts with agency enhancement theory, in my view. If someone has a typical attribution-self-representation system, the person is probably pretty functional, as far as that particular system goes. (There might be cognitive deficits elsewhere, but we can set those aside). If someone is functioning perfectly well, we have no reason to coax her to change her attribution style and self-representation beliefs. If she falls in love or experiences a break-up, we can accept her appraisal of the situation and support her adaptive self-attribution biases in both cases.

But when addressing a person with depression, the fitting response is to be liberal with praise and conservative with blame, inasmuch as people with depression have an excessively pessimistic attribution style and low self-esteem. They already blame themselves more than is healthy. Thus, if a person with depression takes 'too much' responsibility for a break-up, and suffers a loss of resiliency, control, and narrative coherence, we ought to encourage her to 'externalize' the break-up and take more responsibility for positive aspects of the event or relationship.

In addressing narcissistic people, however, the opposite policy is fitting: we should be conservative with praise and liberal with blame, to discourage their excessive egocentrism. We should encourage the narcissist to take more responsibility for the break-up, and less responsibility for positive features of the relationship. On the agency enhancement view, the reason for withholding blame from people with depression and withholding praise from narcissists (relatively to the general population) is that depression and narcissism are antagonistic to adequate

functioning – in particular, to the adaptive functioning of the attribution-self-representation system. Depression is antagonistic because it can, if severe, cause subjective distress, dysfunction, volitional deficits, excessive shame, relationship difficulties, and so on. Narcissism is antagonistic to adaptive functioning because it prevents people from caring about other people's interests and investing in people's projects. Both attribution styles, then, impair agential functioning, albeit in different ways. People with depression might not be able to undertake moral projects or support others' moral projects due to volitional deficits, whereas narcissists may have no interest in engaging in moral projects due to global egoism.

An agency enhancement account recommends that we discourage these extreme tendencies. It counsels that we address people in ways that positively affect their attribution-self-representation system, prodding their biases toward the median. Doing this will enhance their functioning and therefore their ability to pursue projects of value.

It's interesting to note that we do, as a matter of fact, tend to suspend or modify our normal blaming responses toward people with depression. Philosophers have commented on this, and while there is disagreement, there is also meaningful consensus. For example, W. M. Martin argues that "explicit moral language is never appropriate during therapy" (2012: 39), which suggests that full suspension of blame is fitting for psychotherapy service users. Duff Waring counters that it might be helpful to hold service users responsible for certain things, though not in an "excessive, harsh, [or] severe" way (2012: 46). This seems to imply that, while full suspension of the reactive attitudes[4] would be gratuitous, there's good reason to refrain from adopting our normal blaming responses toward people with extremely pessimistic biases and depressive tendencies. Since 'holding responsible' encompasses a plethora of reactive attitudes, including praise and approbation, Duff's view seems right: surely we should not suspend *praise* from all therapy service-users. Martin, then, is wrong if he means to suggest that *every* manner of 'holding responsible' is off-limits in therapy. But it's interesting that Martin and Duff fully agree that we should, at a minimum, attenuate our *blaming* responses toward people with depression, which is consistent with an agency enhancement account, on which blame is sensitive to attribution style. Duff and Martin don't say much about praise *per se*, but, using the pragmatic grounds they offer

---

[4] See Strawson 2008 for an account of the reactive attitudes as central to our moral practice. These attitudes include praise, blame, resentment, gratitude, approbation, and disapprobation.

for withholding blame (i.e., it is counter-therapeutic for some service-users), we can extrapolate that we ought to attenuate our normal *praising* responses toward people with narcissistic traits (e.g., exaggerated self-importance), since their self-regard is excessive. This supposition fits with the agency enhancement theory of responsibility, and with an outcome-based approach to psychological therapy.

Now, I think it's a reasonable conjecture that our reactive attitudes naturally (i.e., implicitly and automatically) respond to people's attribution styles, because they evolved to respond differentially to different expressions of attribution bias (to depression and narcissism, for instance). If so, they would naturally tend to function this way whether we realize it or not. But even if this evolutionary story is wrong, there are compelling pragmatic reasons to think that we *should*, if possible, use our knowledge of attribution theory to try to respond differentially to different attribution-self-representation styles, as this strategy would help us push people in the direction of better functioning. Being functional can help people pursue prosocial (moral) projects and relationships.

6. Gender and self-attribution styles

There are notable gender implications for this theory. There is evidence that narcissism tends to be higher in men than in women (Grijalva et al. 2015), and that women are twice as likely to experience depression as men (Albert 2015). This means that there may be average gender differences in the attribution styles of men and women. As a result, in the absence of any specific biographical information, it may be a good general strategy to attribute less blame to women in the aftermath of a break-up, and more blame to men. More broadly, it may be a good general strategy to give more praise to women and less to men, on average, knowing that women are more prone to depression, and therefore more prone to pessimistic self-appraisals. That said, gender is a defeasible heuristic, which doesn't take into account specific differences between individuals. Still, if we are not familiar with a person's psychological profile, this could be a useful heuristic for deciding how best to hold the person responsible until we become better acquainted with the individual's dispositions.

Self-attribution theory also has implications for the responsibility attributable to men and

women in situations of intimate partner violence (IPV), which affects women more severely than men (with one in three women experiencing severe violence from an intimate partner compared to only one in seven men [NCADV 2011; WHO 2017]). Victims of IPV show higher levels of self-blame, and high self-blame correlates with post-traumatic stress disorder and depression (Levin 2011). Thus, women, who already exhibit higher average levels of depression, are more likely than men to experience severe intimate partner violence, and those with higher levels of self-blame are more likely to experience adverse psychological consequences that could impair their life prospects and future wellbeing.

According to the proposed model of responsibility (which is sensitive to considerations about self-attribution style), women in abusive relationships should be held responsible in such a way as to decrease their depression and enhance their self-efficacy and psychological wellbeing. Rather than encouraging them to identify with the abusive partner (as a neurotypical person would normally identify with a genuine loving partner), we should emphasize that the relationship is *not* part of their deep self, that the abusive partner is *not* a positive part of their self-conception, and that they lack full responsibility in the relationship, but can regain responsibility by leaving it. We should also avoid blaming the person, as blame increases depression, and people in abusive relationships are already high in depression and self-blame. On a therapeutic model of responsibility, it would be preferable to *praise* the person for her resilience and capacity for growth, rather than blaming her for any features of the relationship.[5]

7. Final remarks

I have argued that romantic love exemplifies our tendency to self-attribute positive rather than negative experiences, and I have explained this tendency as an expression of a properly functioning attribution-self-representation system. I have argued that a moderate amount of self-serving bias is healthy, and therefore we should not discourage people from manifesting it. Thus, we should not blame or praise people for moderately over-identifying with positive romantic experiences and moderately externalizing unpleasant break-ups, inasmuch as this attribution bias has adaptive value. I have also shown how depression, or an overly pessimistic attribution style,

---

[5] These are, again, generalizations. Familiarity with a person allows us to make more specific and fine-grained responsibility attributions in light of the person's psychological profile.

and narcissism, or an overly optimistic attribution style, might impair the agent's ability to form romantic bonds and invest emotionally in other people. When there is impairment, we should respond differently to depression and narcissism, so as to positively shape the person's attribution-self-representation system. The moral enhancement theory of responsibility coheres with the view that our interactions with others should regulate their psychological dispositions in ways that improve their agency.

Bibliography

Abramson, L. Y., Selgiman, M. E. P., & Teasedale, J. D. 1978. Learned helplessness in humans: critique and reformulation. Journal of Abnormal Psychology, 78, 40-74.

Albert, P. R. 2015. Why is depression more prevalent in women? *Journal of psychiatry & neuroscience: JPN*, *40*(4), 219.

Badiou, A., & Truong, N. 2012. *In praise of love.* Profile Books.

Bentall, R. P. 2004. *Madness explained: Psychosis and human nature*. Penguin UK.

Brogaard, B. 2015. *On romantic love*. OUP.

De Botton, A. 2016, May 28. *Why you will marry the wrong person. The New York Times.* Retrieved from http://www.nytimes.com/2016/05/29/opinion /sunday/why-you-will-marry-the-wrong-person.html?_r=0.

Fischer, J. M. 2012. *Deep control: Essays on free will and value*. Oxford: Oxford University Press.

Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. 2015. Gender differences in narcissism: A meta-analytic review. *Psychological bulletin*, *141*(2), 261.

Kane, R. 1985. *Free will and values: adaptive mechanisms and strategies of prey and predators*. SUNY Press.

Greenier, K. D., Kernis, M. H., McNamara, C. W., Waschull, S. B., Berry, A. J., Herlocker, C. E., & Abend, T. A. 1999. Individual differences in reactivity to events: re-examining the roles of stabiliity and levels of self-esteem. Journal of Personality, 67, 185-208.

Levine, A. 2011. Intimate partner violence typology, self-blame, depression and PTSD among homeless women.

Martin, W.M. 2012. Psychotherapy as cultivating character. *Philosophy, Psychiatry, & Psychology, 19(1), 37-39.*

McGeer, V. 2013. Civilizing blame. *Blame: Its nature and norms*, 162-188.

Mele, A. 1995. *Autonomous Agents.* New York: Oxford University Press.

Rubenstein, L. M., Freed, R. D., Shapero, B. G., Fauber, R. L., & Alloy, L. B. 2016. Cognitive attributions in depression: Bridging the gap between research and clinical practice. *Journal of psychotherapy integration*, *26*(2), 103.

Shoemaker, D. 2015. Ecumenical attributability. In Clarke, R., McKenna, M., & Smith, A. M.

(Eds.). (pp. 115-140) *The Nature of Moral Responsibility: New Essays*. Oxford University Press, USA.

Strawson, P. F. 2008. *Freedom and resentment and other essays*. Routledge.

Stucke, T. S. 2003. Who's to blame? Narcissism and self-serving attributions following feedback. *European Journal of Personality*, *17*(6), 465-478.

Tennen, H., & Herzenberger, S. 1987. Depression, self-esteem and the absence of self-protective attributional biases. Journal of personality and social psychology, 52, 72-80.

Waring, D. 2012. Psychotherapy through the prism of moral language. *Philosophy, Psychiatry, & Psychology*, 19(1), 45-48.

Vargas, M. 2013. *Building better beings: A theory of moral responsibility*. OUP Oxford.

Vargas, M. 2005. The trouble with tracing. *Midwest studies in philosophy*, *29*(1), 269-291.