I could have been more courageous I will be more forgiving. Such thoughts are frequent, especially at the beginning of the year, like now. We believe that we stand for certain values, or deeper virtues, which might not be expressed or realized in our day-to-day actions. We fall short of a certain idea of ourselves. What Nina et al.'s paper argues is that we tend to transform this shortfall into the idea of a deep true self, distinct from the superficial or more contextual self we express in everyday life. This is the claim I want to focus on, before questioning their second claim that this is not just a self-serving bias.

The original claim of the paper is not that we entertain idealized moral versions of our selves. It is about the way we relate to our better selves. As I acknowledge for instance that I could have been more courageous, or will be more forgiving, I run counterfactuals or projections of a higher moral self. My thoughts then refer to another version of myself, or another temporal slice of myself: I am putting my self in the shoes of Ophelia-if-things-had-been-different, or Ophelia-in-2017.These are other selves, which I temporally substitute to my current self.

But that's not how the true self works. What is clear is that the true self is not a *counterfactual, future* or alternative version of me. The true self is thought to be me: It is actual, it is something which exists at the same time as the superficial self that I express in everyday life. When saying it is actual, I take it that it is not even just thought to be a set of moral dispositions or capacities, but a set of actual, fully possessed virtues. But if we all think that we actually have – or are – a true self, are we all so divided? While the paper makes a clear and convincing case for *what* the true self covers (i.e. moral traits), *how* we exactly relate to this true self, needs more clarity. How is the distance between deep and superficial self managed – and what do people think about the relation between these two 'selves' – if they are indeed thinking about this as two different selves?

I have nothing against the idea that concepts like 'personhood' or 'self' are used flexibly; I have nothing either against attributing inconsistent representations to people: Certainly, we might both think that we are one, and yet two. But one would want an explanation of how this 1-2 selves work, especially when there are good precedents on the philosophy-market. Kant thought that there was a transcendental moral self in us as well as in everyone – is the true self similar, and are we natural Kantians? Harry Frankfurt suggested that we have higher-order desires – a tendency to reflect on the larger values that our more worldly desires and motivations lead us to want and do. Isn't it possible to capture the evidence listed in favour of a true self in this more minimal way : We take some distance from our immediate desires and actions, and wish for that what we want and do to be loftier, nobler, more moral. What would be lost if the contrast between 'true; and 'superficial self' was changed for the contrast between the life we would like to live and the life we live? Other candidates also exist, note, in psychology: As shown by Tali Sharot, we tend to be over-optimistic about our own abilities, and diffuse bad evidence when it contradicts this belief. The beliefs about the 'true self' resemble a general optimism about human morality.

These alternatives raise questions: Couldn't the evidence beautifully reported in the paper be explained by ways of thinking about ourselves other than the dichotomy between true and superficial self? Here is another way to raise the question: Accept that the authors are right, and that we evolved to have such belief into a core, true self. What exactly does the belief in a true self do for us – and more crucially, what does it do that all these other ways of thinking about better moral us (as counterfactual or optimistic versions of us, as dispositions, or through higher-order desires) could not do?

Let me suggest a possible direction of defense: thinking about the true self as actual and part of us might be more optimistic than thinking of counterfactual or future versions of ourselves, or higher-order desires, which are more often accompanied with regrets or sense of striving. Our true self is here already – no need to look further. True self is not something abstract or distant – it is us. So

representing an actual true self might be more useful to us. It is also a much better argument for reputation management: “Look, I can say, this is not really who I am”.

That we extend this optimism about ourselves to others looks like a source of further questions – for the discussion, hopefully!