Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-025-10483-2

Diachronic or Counterfactual? Temporal Well-Being and Changing Attitudes

Marina Moreno[1]

Accepted: 31 January 2025

Abstract
In his paper "Wellbeing and Changing Attitudes Across Time", Krister Bykvist investigates the challenge that changing attitudes pose for attitude-sensitive theories of well-being in determining temporal wellbeing. He offers both a useful tool to investigate the conceptual space of possible answers, namely an attitudinal matrix, as well as a seemingly plausible constraint on such an answer, which he terms 'diagonalism'. This paper draws on his matrix to provide an argument against diagonalism and offer an error theory regarding its intuitive force. Using his matrix, I sketch my own view of temporal well-being according to which it is not a single unified concept, but can rather be understood in two different ways ('better for' vs. 'better off'), both of which are interesting in their own right.

Keywords Attitude-sensitive well-being · Temporal well-being · Counterfactual attitudes · Diachronic attitudes

1 Introduction

Attitude-sensitive theories of well-being hold that a person's well-being is, at least in part, determined by their attitudes. In his paper "Wellbeing and Changing Attitudes Across Time", Krister Bykvist investigates the challenge that changing attitudes pose for such attitude-sensitive theories of well-being. If attitudes about periods of your lives change, it seems as though attitude-sensitive theories of well-being cannot provide a stable answer as to how much temporal well-being a particular period contains. Temporal well-being is to be understood as separate from lifetime well-being: The former refers to well-being of a particular time while the latter refers to the total well-being of a whole lifetime. If attitude-sensitive theories of well-being cannot provide a stable evaluation of temporal well-being, this constitutes a significant drawback for such theories (431). Bykvist takes up this challenge. In this comment, I employ the helpful attitudinal matrix he introduces to argue against his solution. I argue that temporal well-being is not a unified concept and must be

✉ Marina Moreno
marinaestrellamoreno@gmail.com

[1] Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität Munich, Ludwigstr. 31, 80539 Munich, Germany

Published online: 15 February 2025

disambiguated into seperate questions. And I offer an error theory regarding the intuitive force of Bykvist's objections against such a view.

2 Attitudinal Matrix and Three Possible Answers

Thomas Nagel offers the following example to illustrate the problem of changing attitudes: In your youth, you favoured[1] an adventurous over a quiet lifestyle, while, in your old age, you favour a quiet lifestyle over an adventurous one (Nagel 1970, p. 70). Now, say you live an adventurous life when you are young and a quiet life when you are old - what is the temporal well-being for each of these times? To simplify the discussion, let us represent the attitudes by utilities: Say your young self's utilities are $+10$ for the adventurous life and -5 for the quiet life, while your old self's utilities are -3 for the adventurous life and $+7$ for the quiet life.[2] Bykvist introduces a very helpful two-by-two matrix of the following sort to illustrate respective situations (cf. Bykvist 2022, 4):

Case 1

	Period 0	Period 1
Young Self	10	-5
Old Self	-3	7

Bykvist makes use of this matrix to distinguish three different versions of attitude satisfaction theories based on how they evaluate temporal well-being: The attitude theory, the object theory and the satisfaction theory. I am skeptical as to whether these theories do actually evaluate temporal well-being in the ways Bykvist lines out. This is partly because, as I argue later, these different ways of evaluating temporal well-being do not seem to be mutually exclusive, while the theories Bykvist claims they stem from are. Given the scope of this comment, I cannot go into this skepticism any further. To sidestep this issue, I will thus use the name of these theories to refer to the ways Bykvist claims they evaluate temporal well-being, rather than to the substantial theories he claims they stem from themselves.

According to the attitude theory, p_0 (period 0) is to be evaluated given the attitudes you have during p_0 towards both p_0 and p_1 (period 1). That is, the utilities for p_0 are evaluated given the horizontal utilities of your young self. In the above case, this would mean that the temporal well-being of p_0 is $10-5=5$. Vice versa, the temporal well-being of p_1 would be $-3+7=4$. The object theory, on the other hand, evaluates periods of time according to all the attitudes you have at any given time towards that period. For instance, the temporal wellbeing of p_0 is evaluated given both the attitude of your young self towards p_0 and

[1] I will refer to attitudes with various terms such as "favouring", "valuing", etc. This is not to be understood as a stance on what attitudes ultimately are or which ones matter, but merely a convenient way of speaking. The reader may insert their own preferred understanding of attitudes.

[2] Note that my talk of different "selves" is not meant to indicate any metaphysically controversial position about personal identity or the nature of persons in general. Referring to "selves" is merely a convenient way of distinguishing versions of a person who hold different sets of attitudes at different times in their lives. The boundaries of these "selves" are pragmatically given by however long the relevant set of values persisted. Following Bykvist, I will always assume that all of these periods are equally long.

your old self towards p_0. In the above case, the temporal well-being given by p_0 would be $10-3=7$, while the temporal well-being given by p_1 would be $-5+7=2$. Lastly, the satisfaction theory discounts diachronic attitudes and holds that only the diagonal utilities matter, i.e., in our example the temporal well-being given by p_0 would be 10 and the temporal well-being given by p_1 would be 7.

Bykvist endorses the satisfaction theory and argues for a constraint for theories which he terms "diagonalism", i.e., that the attitudes that are relevant to evaluate temporal well-being are exclusively represented by the top left and bottom right cell of the attitudinal matrix. This view, he argues, has several virtues: It provides a non-relative standard to evaluate temporal well-being, satisfies the resonance constraint, according to which a time T is good for you only if it resonates with you *at* T (cf. Bradley 2016 for a discussion of a version of this constraint), and it avoids important counterexamples he sees both the attitude theory and object theory confronted with (Bykvist 2022, 7, 13). Before I go into these counterexamples, however, it is important to first get a better grasp of the kind of question we are asking to begin with.

3 What is the Question of Temporal Well-Being?

Which question are we really asking when we talk of the problem of temporal well-being given changing attitudes? I believe that Bykvist himself does not make sufficiently clear which question he is interested in precisely. In clarifying the challenge he is investigating, he goes over Nagel's example and notes the following: "But this would not provide us with a stable answer to the question of which time is better for you! This means that there is no stable answer to whether things are becoming better or worse for you, and we can't say whether you are better off or worse off now than in the past." (431) This quote alone seems to asks at least two, possibly different questions:

(i) Which time period is better or worse for you?
(ii) Are you better or worse off in a particular time period?

Of course, these questions look like they are at least closely related, but it is not clear from the outset that they are identical. Indeed, I will argue that we can ask two semantically distinct questions, which roughly correspond to the two questions above, and that "temporal well-being" is thus not a unified single concept.

To see this, I propose to simplify our problem first by removing one set of temporal indices. Say you are gifted a watch and a ring when you are young, but keep both of these objects throughout your life. However, your attitudes towards the objects change over time. We can, again, use a two-by-two matrix to depict a possible such situation:

Case 2.

	Watch	Ring
Young Self	10	-5
Old Self	-3	7

In this example, your young self values the watch a great deal but dislikes the ring, while your old self values the ring a lot but dislikes the watch. There is an important insight we can draw from this example, namely that we can ask two semantically different questions about how the watch and the ring contribute to your well-being. First, we can ask how the ring and the watch jointly contribute to your well-being while you are young (or old), i.e., how well your attitudes towards both the ring and the watch are satisfied when you are young. This seems to be answered by an analogue of the attitude theory (i.e., adding the horizontal utilities). Second, we can ask how the ring (or the watch) contributes to your well-being throughout your life, i.e., how good the ring as an individual object is for you overall. This seems to be answered by an analogue of the object theory (i.e., adding the vertical utilities). Both of these questions seem interesting and legitimate in their own rights, rather than being mutually exclusive.

With the ring and the watch, we can draw a clear distinction between the objects of an attitude and the time at which the attitude is held, which highlights the fact that the attitude and the object theory do not seem to be mutually exclusive. But this difference is obscured when the time periods themselves become the objects of the attitudes. It then turns out that time periods are two things at once: They are both the object of a person's attitude as well as the time at which a person holds attitudes. These two different roles is what makes temporal well-being an ambiguous concept in the context of attitude-sensitive theories of well-being: We can break it up into at least two different questions.

We can ask, first, how well a person's joint attitudes are satisfied *during* a particular time period. Answering this question will have to take into account all the attitudes the person holds at the time, including the attitudes about all the time periods that are the objects of these attitudes. This first question thus seems to be answered by the attitude theory. And it seems, roughly, to correspond to how well off a person is *during that time*, i.e., to correspond to the above question (ii). After all, we are asking how well the person is faring at the time, i.e., according to an attitude-sensitive theory of well-being, how well what they care about is satisfied at the time.

But time periods can also be the object of a person's attitudes. We can thus, second, ask: How good of an object is the particular time period for the person, given all of their attitudes towards that object? This second question seems to be answered by the object theory. And it seems, roughly, to correspond to how *good a time period* (as an object) is for a person, i.e., to correspond to the above question (i). After all we are asking how the time period as such measures up with respect to the well-being of a person, i.e., according to an attitude-sensitive theory of well-being how well that time period satisfies (all of) the person's attitudes.[3]

We have thus identified two different questions that are each answered by a different theory. Interestingly, neither of them is answered by the satisfaction theory, despite this being Bykvist's favoured theory when it comes to temporal well-being. Let us thus investigate the objections against the object and attitude theories.

[3] I thank an anonymous reviewer for the helpful suggestion to understand the two different concepts of temporal well-being as a difference between 'better off' and 'better for'. Yet I do not ultimately insist on that the difference between these two concepts of temporal well-being are precisely captured by the better off / better for distinction. My main point is that there are two distinct concepts here which need to be separated but are both legitimate in their own right. It seems to me that they correspond to something like the difference between being better off and better for, but if we can find other terms which better capture this difference, I would be just as happy, i.e. my attitudes would be just as satsified.

4 An Error Theory for Diagonalism

Bykvist provides a counterexample for both the object theory and the attitude theory. The counterexample for the object theory is illustrated by the following matrix (Bykvist 2022, 13):

Case 3.

	Period 0	Period 1
Young Self	-2	-4
Old Self	2	4

According to the object theory, both the temporal well-being of p_0 as well as p_1 would be 0 ($-2 + 2$ and $-4 + 4$). However, if we look at the well-being diagonally, it seems that the life clearly improved, given that the young self disvalued their own time, while the older self values their own time. Bykvist thus argues that the object theory cannot account for the intuition that such a life would clearly have improved. The counterexample for the attitude theory works analogously:

Case 4.

	Period 0	Period 1
Young Self	-2	2
Old Self	-4	4

Here again, the temporal well-being of both p_0 and p_1 is 0 if we adopt the attitude theory, given that we now sum horizontally. However, Bykvist likewise insists that the respective life has clearly improved, i.e., that p_1 is better than p_0. But the attitude theory cannot capture this intuition. Since both the object theory and the attitude theory cannot capture such clear intuitions, Bykvist argues, the satisfaction theory has a strong advantage over them (Bykvist 2022, 12ff). However, in what follows, I will argue that these intuitions stem partially from an intuitively false understanding of such cases and that the disambiguated concept of temporal well-being can capture the remaining intuitive force of the examples.

To understand my response, take the above example with the watch and the ring again and let us reintroduce a second set of temporal indices. Say that the young self is gifted the watch but loses it, while the ring is gifted only later to the old self. Introducing these temporal indices means that we must reinterpret the diachronic utilities. For in what sense does the young self disvalue the ring, if they do not themselves possess it in the first place (and vice versa for the old self)? There are at least two possibilities. The first possibility is that the young self values the ring in a counterfactual sense, i.e., they *would* have a utility of -5 *if* they had the ring. But given that they do not in fact have it, the ring does not actually have an impact on their attitude-satisfaction. The second possibility is that their attitude is not about the ring itself, but about the fact that the old self has a ring. In this case, the fact that the old self has the ring does indeed have an impact on the attitude-satisfaction of the young self.

In the former case, the attitudes of the young self and the old self do not, in fact, diverge over the same object: The young self has an attitude about the (nonactual) state of the young self owning a ring, while the old self has an attitude about the (actual) state of the old self owning a ring. In such a situation, they might not actually disagree at all. That is, the young self might value *for the old self* to own a ring, and the old self might value *for the young self* not to own a ring. In order for there to be a genuine conflict between their diachronically changing attitudes, then, the utilities have to be interpreted in the latter sense, i.e., the young self disvalues in a direct manner that the old self owns a ring. The following matrix captures this situation:

Case 5.

	Owning a watch in p_0	Owning a ring in p_1
Young Self	10	-5
Old Self	-3	7

Now the satisfaction theory holds that what is relevant for your temporal well-being in p_0 is merely the fact that you own a watch, and not the fact that you will own a ring, while what is relevant for your temporal well-being in p_1 is merely the fact that you own a ring, and not the fact that you used to own a watch. Spelled out in this way, however, I fail to see the appeal of the satisfaction theory as a theory of temporal well-being. We just postulated that the fact that the old self owns a ring does in fact have an impact on the young self's attitude satisfaction, so why are we now arbitrarily excluding this impact from the question of how well off they are at that time? If the young self genuinely disvalues the fact that the old self will own a ring, why should we discount this attitude, just because whether or not it is satisfied depends on a fact outside of the temporal extension of the attitude (cf. Dorsey 2018 for a similar argument regarding benefits conferred on past selves)?

If this is correct, the satisfaction view cannot be upheld. I believe that its intuitive force comes from underlying confusion about many classic examples, including Nagel's example. We may, contingently, often have merely counterfactual attitudes (as outlined above) regarding times outside of our temporal extension. For instance, say that I loves grapes, but I know my old self will come to dislike them. If my young self has an attitude towards having grapes when I am old, such an attitude is most likely going to be counterfactual: I do not actually care about whether *my old self has grapes*. When I evaluate having grapes as positive, I likely do so because *If I had the grapes right now*, I would value that.[4] Our intuitions might thus rest on such counterfactual attitudes, even though we might not be consciously aware of it, given that we rarely make this distinction explicit. However, if the disagreement between two selves is merely counterfactual, it is clear that diagonalism is very intuitive: If both the young and old self are happy to defer to each other regarding how they spend their "own" time, the diagonal values are clearly the only relevant ones to evaluate actual well-being.

This error theory in combination with having disambiguated the concept of temporal well-being into two distinct questions provides us with resources to defend against

[4] I thank an anonoymous reviewer for this illustrative example.

Bykvist's counterexamples. Consider first the counterexample against the attitude theory. He argues that it is implausible to hold that someone's life does not improve if their synchronic attitudes improve. That is, if your young self disfavours their own time and favours the future and your old self favours their own time and disfavours the past, it seems that your life must have improved. First, even if we say, in accordance with the attitude theory, that *you are equally well off* in both times, we can still, in accordance with the object theory also at the same time say that *the second period is better for you*. We can thus capture some intuitive sense in which the life improves.

Second, if the intuition persists that *you really are better off* (and it is not merely that the time is better for you), this may well be due to a similar intuitive confusion regarding actual and counterfactual attitudes as outlined above. Again, if the diachronic attitudes of each of the selves are merely counterfactual, it is intuitively plausible that the life improved, since only the diagonal attitudes are actual (rather than counterfactual).[5] But if their diachronic attitudes are actual, the time periods at which they do not exist nevertheless generate genuine satisfaction or dissatisfaction for them. Indeed, it seems that both your young and old self would actually *agree* that you are overall not better off. The young self might say "My own time was not very valuable, but the knowledge and hope of a better period of time in the future made life worth living for me.", while the old self might say "My own time was generally good, but the knowledge of my past struggles generally made life less valuable for me, such that it was just worth living."

An analogous argument can be constructed against the counterexample to the object theory. Bykvist argues that if your young self dislikes both their own time and the future, and your old self likes both the past and their own time, your life must clearly have improved. Now while we, in accordance with the object theory, can say that both time periods *are equally good (objects) for you*, we can still say that *you are better off* when you are older. We can thus, again, capture some sense in which your life has improved. And, again, any remaining intuitions may well be due to an intuitive confusion about counterfactual versus actual attitudes.

5 Complexity and Well-Being

A final objection before I conclude: If temporal well-being is not a unified concept but has distinct questions associated with it, on the basis of which of these distinct concepts should we make decisions? Does this not introduce unnecessary complexity as to how different times relate to our well-being? Bykvist might insist that his theory is simpler, since it always gives one clear answer when temporal well-being is concerned. Note that Bykvist concedes that diachronic attitudes may have a certain kind of atemporal value, which are relevant for the evaluation of one's whole life, but not for any particular temporal well-being (Bykvist 2022, 13–14). While he can thus accomodate the intuition

[5] Empirically, our attitudes to, e.g., our past are likely usually not entirely counterfactual: People often have genuine regret for their difficult pasts. Yet, it is likely not very prevalent that this regret is *equally strong* as the aversion was when this difficult past was actually experienced. An equally strong aversion will likely only be found in the counterfactual evaluation of such a state, i.e. in evaluating it as if it were real in the present. This is, as I argue, where the intuitive confusion comes from.

that diachronic attitudes should count for something too, his view has some considerable complexity too. Yet, it may still be said that several different understandings of temporal well-being do worse on this account.

I cannot do this objection full justice here. But let me sketch two brief answers. First, on my view, both ways of understanding temporal well-being always sum to the same amount of lifetime well-being. The relationship between temporal well-being and lifetime well-being is thus, compared to Bykvist, at least quite straightforward. Second I do believe that disambiguating the concept of temporal well-being into two distinct concepts is useful for decision-making purposes, since based on what decision one faces exactly, a different notion of temporal well-being might be relevant. Due to space constraints, I can only sketch two brief examples to illustrate this.

Say we must decide whether to create a person with life A or life B. The two lives are identical, except that for life B, there is one additional value change (i.e. an additional 'self') and associated time-period, towards which all other selves have diachronic attitudes too. Whether or not we add this period to the person's life plausibly depends on whether the additional period is good or bad for them, i.e., how the additional time period fares as an object of the person's attitudes. This suggests that whether we should choose life B over life A depends on the object theory's evaluation of it, i.e., whether this additional time period is *good for the person*. Put differently: The difference in overall lifetime well-being between life A and life B is precisely the difference the object theory captures.

But say I can choose which time-period of their life I want to spend with someone, and I want to make sure I spend the time with them when they are worst off, such that I can be most comforting to them at the time. Locating the time period at which they are worst off then seems to depend on when the attitudes they hold are least satisfied. After all, whether some attitudes are satisfied that a self does not currently hold cannot affect the well-being they experience at the time. Put differently: The particular part of the overall lifetime well-being that is experienced at a particular time period is precisely what the attitude theory captures. For this decision, then, the attitude theory's evaluation would be relevant, i.e., at which point in time the person is worst off.

If this is right, disambiguating the concept of temporal well-being is thus practically important and does not add complexity for no reason. However, how exactly one would have to take the differing attitudes of the differing time periods into account will likely depend on more than just how we understand temporal well-being. For instance, Pettigrew (2019) argues at length that factors such as how much we identify with the values of other times will play a role too. Nevertheless, correctly understanding temporal well-being is an important part of this puzzle.

6 Conclusion

I have argued that the challenge of changing attitudes reveals that the concept of temporal well-being needs to be disambiguated into separate questions and concepts. While the attitude theory captures how well off someone is at a certain time period, the object theory captuers how good a time period is for someone (overall). I further argued that the intuitions behind Bykvist's counterexamples against both the attitude theory and the object theory are partially based on an intuitive mistake about actual versus counterfactal attitudes.

Acknowledgements I am grateful to Christian List, Hein Duijf, Johanna Thoma, Felix Lambrecht and two anonymous reviewers for feedback and comments on the ideas developed in this paper.

Author Contributions The author has sole authorship of this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Janggen-Pöhn-Stiftung.

Data Availability Not applicable.

Declarations

Ethical Approval Not applicable.

Informed Consent Not applicable.

Competing Interests The author has no relevant competing financial or non-financial interests to disclose.

Statement Regarding Research Involving Human Participants and/or Animals Not applicable.

References

Bradley B (2016) Wellbeing at a time. Philosophic Exch 45(1):1–12
Bykvist K, Practice M (2022) Wellbeing and changing attitudes across time. Ethical Theory Moral Pract 1–15. https://doi.org/10.1007/s10677-022-10311-x
Dorsey D (2018) Prudence and Past selves. Philos Stud 175(8):1901–1925
Nagel T (1970) The possibility of Altruism. Oxford Clarendon Press
Pettigrew R (2019) Choosing for changing selves. Oxford University Press, Oxford, UK