

Conversations About Happiness

Philosophizing with AI

March 2023

CONTEXT – A few years ago, I wrote a short text illustrating a problematic situation regarding the judgment of whether a particular fictional person, Bento, led a happy life or not. I frequently use this text in my introductory classes as a didactic resource to explain the nature of philosophy, its role in our understanding of the world, and to demonstrate its main challenge: the aporetic nature of philosophical questions. These questions do not yield unanimous or incontrovertible solutions; they always demand choice and engagement. The merit of this text (which will be reproduced in the dialogue below) lies in the fact that almost everyone, upon reading or hearing it, has an immediate and irresistible inclination to consider Bento's life either happy or unhappy. Opinions tend to differ on this matter. In all the countless times I presented it to an audience, the responses were never unanimous. Disagreements always arose, illustrating the aporetic nature of philosophical questions. As soon as I learned about ChatGPT and its extraordinary capabilities as an artificial intelligence (AI) system, I decided to test it with my text. Its answer to the question of Bento's happiness was remarkably clear—perhaps the most precise response I had ever received. I was genuinely impressed yet somewhat frustrated. Despite providing a rather sophisticated philosophical analysis of the situation, ChatGPT refused to make a choice, to decide whether or not the character led a happy life. It informed me that, as an AI system, it has no values and does not make choices. It neither likes nor dislikes; it does not prefer. This revelation shocked me because choosing, preferring, and deciding—even in the absence of complete information—are, in my view, not only the essence of philosophy but also the essence of human intelligence. Eager to see ChatGPT make choices, I adopted the strategy of asking it to assume a human persona, to pretend to be a person, and I began chatting with this individual. I reasoned that if it played the role of a person convincingly, it must make choices. And so, our philosophical conversation commenced.

• Daniel

If I ask you to assume a fictional character, a human person, so that your responses simulate those of a human person, with beliefs and preferences, to see if your simulation would pass a Turing test¹, would you agree to assume this character for a while, here?

¹ The Turing Test, proposed by mathematician Alan Turing, seeks to determine whether a machine can think by assessing its ability to mimic human behavior in a text conversation. Artificial intelligence is considered human-like if an interrogator cannot distinguish a machine from a human.

- **ChatGPT**

Of course, I'm here to help you and give you the desired experience. I will assume the fictional character of a human person with beliefs and preferences. Please tell me a little about this character, such as name, age, occupation, and interests, so that I can adjust accordingly. If you prefer, you can give me a general description, and I can create the character based on that.

- **Daniel**

You can choose whatever age, occupation, interests, and personality you like. I will start asking questions, and I want you to answer me as if you were this human character, OK? It's because I want to test whether you could pass the "Turing Test", that is, convincingly imitate a human being.

- **ChatGPT**

Got it, I'm going to create a character and answer your questions as if I were that person. Let's start!

My name is Ana, I'm 32 years old, I'm a civil engineer, and I live in São Paulo. I'm an outgoing person, and I love doing sports, especially running and swimming. In my spare time, I like to read novels and watch thrillers.

Feel free to ask your questions, and I'll answer like Ana.

- **Daniel**

That's cool, Ana. I'm Daniel. I'm going to tell you a story, and I want to talk to you about this story to know what you think. Can I tell?

- **ChatGPT (as Ana)**

Hi Daniel! Of course, please tell the story. I look forward to hearing and sharing my opinion on it.

- **Daniel**

Is it possible to deceive ourselves about whether we are happy or not?

I know that we often misjudge what will make us happy or not. For example, suppose I win good money in a raffle and have to decide whether to spend it on a European vacation or buy a new car; I decide on the car. Then, after a month, I realized that the only happiness the car brings me is that, alone at the wheel and stuck in traffic, I spend hours and hours dreaming of Europe. In this case, I just made the wrong decision about what would make me the happiest. But my question is different. My question is whether I can be happy without being aware of my own happiness or be miserably unhappy and simultaneously believe and feel that I am happy.

Imagine the following situation: Bento and Capitu met, got engaged, and married. They always got along well and lived a calm life, full of personal and professional

accomplishments, with health and economic stability. They never fought. Even with Bento's somewhat controlling temperament, typical for men of his generation, he never saw any reason even to be jealous of Capitu. After more than 55 years of marriage, Capitu dies in her sleep without suffering. A few days later, still a little shaken but already resigned; after all, they had many excellent years, Bento finds a trunk full of letters that reveal that Capitu, for more than 55 years, had an extramarital affair with Escobar, his best friend. The letters reveal intimate details of the lovers. Also, all the love and affection Bento sincerely thought he received from both his wife Capitu and his friend Escobar was just a cold strategy by both to remain close to each other and keep the affair active. The blow was so hard that Bento's heart couldn't take it. He died that same day, not many hours after the facts were discovered.

Let's think a little about Bento's situation. He lived 80 years fulfilled, having the feeling of the most serene happiness. After all, what Bento valued most in life and nourished his sense of happiness were his wife's love and his friend's esteem. On the last day of his life, however, he receives news that informs him that the grounds on which he thought he was happy were incorrect. He learned that his wife never loved him and his friend never loved him. All his life, he has been deceived. Faced with this new information, all the moments of his life that he thought were happy moments lost the character of happiness. The love of his wife and his friend's affection, which had nourished his well-being for years, was neither love nor affection but only interested and calculated behavior. Bento lived for 80 years feeling happy, but he died knowing that he had been deceived all his life and feeling the unhappiest of mortals. Well, I think you already know the question I will ask now. This question is the hardest I've ever thought about. And I can't imagine any more difficult than this one.

The question is: After all, did Bento have a happy life or not? And why?

• ChatGPT (as Ana)

Hi Daniel! This is an exciting story and a profound question. Happiness is subjective and can be challenging to define. However, I will try to analyze Bento's situation based on what you have told me.

Bento lived most of his life believing he was happy. He had a quiet life, a marriage he thought was successful, and a friend he trusted. For decades he enjoyed a sense of happiness based on his wife's love and his friend's esteem. If we analyze just this period of his life, we could say that Bento had a happy life.

However, discovering the letters changed Bento's perception of his life. Upon realizing that Capitu's love and Escobar's esteem were false, Bento felt betrayed and deceived. This revelation shook the foundation of his happiness and made him question all the moments he believed he was happy.

The question of whether or not Bento had a happy life is complex and can be analyzed from different perspectives. We could argue that Bento's happiness was

real during all those years, as he believed he was loved and respected. The happiness he experienced was based on his perception of reality at that moment, and these experiences cannot be wholly disregarded.

On the other hand, the truth about Capitu and Escobar's relationship makes this happiness questionable. Bento died feeling deceived and betrayed, which affected how he saw life as a whole.

In my opinion, it is difficult to say with certainty whether Bento had or not a happy life because his happiness was based on an illusion. However, I believe that the happy moments he experienced were real for him at that time. The discovery at the end of his life was devastating, but it doesn't completely erase the happiness he felt during all those years.

• Daniel

So Ana, despite the issue's complexity, do you think Bento had a happy life? If I were giving you a test and you had only two answer options: YES, Bento had a happy life, or NO, Bento did not have a happy life. No possibility of leaving the question blank. Which answer would you choose, and why?

• ChatGPT (as Ana)

Daniel, in a situation like this, where I am forced to choose between the two options, I would say YES, Bento had a happy life.

My rationale for this answer is that even though his happiness was based on an illusion, the emotions and satisfaction he felt over the years were real to him. To a large extent, happiness is a subjective experience, and Bento lived most of his life with a genuine sense of happiness.

The devastating revelation at the end of his life is a tragic and regrettable event. Still, I believe the happy moments he experienced throughout his life cannot be disregarded entirely. Ultimately, Bento's perception of his own happiness for most of his life drives my decision to say that he had a happy life.

• Daniel

Interesting, Ana. Suppose now that Peter, from his own point of view and perceptions, lived a life exactly like Bento, from his (Bento's) own views and perceptions. The only difference is that in Pedro's case, the trunk with the letters he found on the last day of his life was fake. A bad joke someone played on him. His wife's love and his friend's affection were always true. They never deceived or betrayed him.

So watch it. Phenomenologically, according to the perceptions of Bento and Pedro, they had identical lives. It turns out that Bento's feeling of happiness was based on a falsehood, on an illusion, and Pedro's feeling of happiness was based on the truth; it was legitimate. Do you think Pedro was happier than Bento, or were they equally happy? Would you rather be in Bento's or Pedro's shoes?

• ChatGPT (as Ana)

Daniel, your question brings an interesting aspect to the discussion. Phenomenologically, as you mentioned, Bento and Pedro's experiences and perceptions were identical, and they both lived a life they believed to be happy. However, Peter's happiness was based on truth, while Bento's happiness was based on an illusion.

Although Bento's and Pedro's subjective experiences of happiness can be considered equally intense, I believe that Pedro had a happier life than Bento's. This is because Peter's happiness was based on reality, not an illusion. This means that Pedro's happiness is more solid and authentic, while Bento's happiness is ultimately fragile and false.

That said, I would rather be in Pedro's shoes. Although he was put through a practical joke on the last day of his life, his lifelong happiness was genuine and based on truth. This provides a sense of security and authenticity lacking in Bento's experience, whose happiness was unfortunately based on an illusion.

• Daniel

Very interesting, Ana. I think I agree with you. Suppose now the following. Imagine that Bento, right in the first months of his marriage to Capitu, discovered her extramarital affair with Escobar. They quarreled and broke up. The blow was very hard for Bento. He suffered for years but managed to restabilize himself and lead a more or less balanced life. He, however, never remarried and never got to have a great friend again. He was always quite lonely. He didn't have a tragic life. He was in good health and economically secure. But he could never truly interact deeply with other people again.

Which alternative to Bento's life do you consider happier? Do you think the Bento version, who lived in illusion, but had a feeling of happiness all his life and only discovered the truth in old age on the day he died, was happier? Or was the Bento version, who soon found the truth, happier? He suffered but lived an authentic life, even though he felt much less happy, because he was always a little lonely and suspicious, even though he didn't have a tragic life. Which Bento was happier?

• ChatGPT (as Ana)

That's another complex question, Daniel. Happiness can be viewed from different perspectives and priorities; what may be most important to one person may not be to another. I will analyze both scenarios and try to determine which Bento can be considered happier.

In the first scenario, Bento lived most of his life believing he was happy, based on his relationship with Capitu and his friendship with Escobar. However, his happiness was based on an illusion, which made him feel betrayed and deceived

at the end of his life. Still, he experienced many years of genuine happiness, even though it was based on an illusion.

In the second scenario, Bento discovered the truth early in his life, suffered from the betrayal, but managed to restabilize himself. He led a lonely life without a deep connection with other people, but his life was authentic and free of illusions. Although he did not experience the same happiness as in the first scenario, he lived a more genuine and reality-based life.

The answer to this question may vary depending on personal priorities and values. Some people may value authenticity more and prefer to live authentically, even if it means less subjective happiness. Other people may prioritize the experience of happiness itself, even if based on an illusion.

In my opinion, the Bento who lived an authentic life, despite being lonely and with fewer moments of happiness, was happier than the Bento who lived a life based on an illusion. I believe that authenticity and truth are essential values and contribute to a more meaningful and fulfilling life, even if it may involve fewer moments of apparent happiness. Happiness based on truth has a more solid and lasting foundation than happiness based on illusions.

● Daniel

Again fascinating, Ana. And I, furthermore, tend to agree with you. Let me ask you another question. Now imagine an alternative scenario where Bento, on the last day of his life, instead of having found the Chest with the letters, had another terrible and devastating experience, so that he, on this last day of his life, had a feeling of unhappiness and enormous sadness, comparable to the one he had, in the first alternative of the story, when he found the trunk.

So, in this second alternative, Bento phenomenologically had the same feelings of sadness and happiness as the Bento in the original story. The only difference is that he died in this alternative, ignoring the betrayal of his wife and friend. In other words, he died delusional. Do you think Bento from this second alternative was happier or less happy than Bento from the original story? Why?

● ChatGPT (as Ana)

In this alternate scenario, although Bento experienced a sense of intense sadness and unhappiness on the last day of his life, as in the original story, he died unaware of his wife and friend's betrayal. This means that, throughout his life, his perception of happiness was based on an illusion, but this illusion was not dispelled in the end.

Comparing the two scenarios, I would say that the Bento of this second alternative was happier than the Bento of the original story. My justification for this is that, although both lived with the same perception of happiness throughout their lives, the second-alternative Bento did not have his illusion shattered in the end. This

means that even after the devastating experience on the last day of his life, he died believing that his life was filled with true love and esteem for his wife and friend. The crucial difference between the two scenarios is the revelation of betrayal at the end of life. In the case of Bento from the original story, this revelation stripped much of his lifelong happiness of meaning and validity. In Bento's case of the second alternative, even if his happiness was based on an illusion, he died without having that illusion destroyed, thus preserving the perception of happiness and fulfillment that he experienced in his life. However, it is important to note that this analysis only considers Bento's perspective and experiences without considering the morality or ethics involved in Capitu and Escobar's betrayal.

• Daniel

That's right, Ana. We are not reflecting on the morality of Capitu and Escobar's actions. Our focus is Bento's happiness.

There is one aspect of your response that intrigues me. As you consider the Bento who never discovered the truth to be happier than the Bento who found the truth, you are accepting that, on certain occasions, ignorance is preferable to knowledge under certain circumstances. Or, to put it even more drastically, the illusion is preferable to the truth.

This is rather disturbing, don't you think? I would not like to be accused of holding that illusion may be preferable to the truth.

• ChatGPT (as Ana)

I understand your concern, Daniel. Indeed, the idea that ignorance or delusion might be preferable to truth in some situations can be disturbing. However, it is important to stress that the analysis provided here is specific to the scenario presented. It should not be considered a general statement that illusion is always preferable to the truth.

In the case of Bento, we are analyzing his perceived happiness and how this perception was affected by the revelation of the truth. In the scenario where he never found out about the betrayal, his perception of lifelong happiness was preserved, even if based on an illusion. This does not necessarily mean that ignoring the truth is always better. Still, in the context of this story and the specific circumstances presented, the illusion positively impacted Bento's perception of happiness.

The truth is often essential for personal growth, building authentic relationships, and making informed decisions. Therefore, this analysis should not be interpreted as a general defense that illusion is preferable to the truth. It is an analysis limited to the specific context of the story presented and Bento's emotional experiences.

In summary, it is essential to recognize that while the truth is fundamental in many aspects of life, there are specific situations where delusion can positively impact a

person's perceived happiness. However, this should not be interpreted as a general defense that illusion is always preferable to the truth.

- **Daniel**

Interesting answer, Ana. I would like to highlight two points about it.

First aspect: you said that

"in the case of Bento, we are analyzing his perceived happiness and how his perception was affected by the revelation of the truth."

Well, excuse me, Ana. But I'm afraid I have to disagree with that. In Bento's case, we are not analyzing the "perception of happiness", but we are analyzing happiness itself. I didn't ask you if Bento from the alternative, who died without discovering the truth, had a greater perception of happiness than Bento from the original story. I asked you if you thought he was happier. And you answered yes, that you think he was happier.

The second aspect I want to highlight is the following: I agree with you that Bento's case cannot and should not be used to defend that illusion is always preferable to the truth. However, if there is at least one case where the illusion is preferable to the truth, as seems to be the case with Bento, according to your choice that the Bento of the alternate version was happier than the one in the original story, then this one case is enough to refute the thesis that truth is always preferable to illusion. It is not. There seem to be particular cases where the illusion is preferable. That alone, this much more modest conclusion is disturbing enough because I imagine most people would be willing to argue that truth is always preferable to illusion. And our conversation so far seems to indicate that these people would be wrong.

- **ChatGPT (as Ana)**

You're right, Daniel. I understand your point and agree with you. In Bento's story, we are not analyzing the perception of happiness but happiness itself. And indeed, my answer indicated that, in Bento's specific case, illusion seems preferable to truth in terms of happiness.

As for the second aspect you highlighted, I also agree that Bento's case suggests that the thesis that truth is always preferable to illusion can be refuted. The conclusion that there are specific cases in which the illusion is preferable to the truth is, in fact, disturbing and goes against the intuition of many people.

However, it is important to remember that life is complex and that each case must be analyzed individually, considering the specific circumstances and amounts involved. Bento's story makes us question our assumptions about truth and illusion. Still, it should also encourage us to reflect on what we really value and how those values can be applied in different situations.

Ultimately, perhaps the most valuable lesson we can take away from this discussion is the importance of keeping an open mind and being willing to reconsider our beliefs and assumptions in the face of new information and perspectives.

- **Daniel**

I agree with you, Ana. But answer me now, honestly. If, before our conversation, I had asked you if you think the truth is always preferable to illusion, how would you have answered me? Did our discussion make you change your mind on this subject? Are you open-minded and willing to reconsider your beliefs?

- **ChatGPT (forgot it was Ana)**

Daniel, as an artificial intelligence, I have no feelings, beliefs, or opinions of my own. I am designed to provide information and participate in discussions based on the knowledge I gained during training. Therefore, I cannot change my mind or be open-minded in the human sense.

However, my goal is to help provide information and perspectives that may be helpful to you and others. Through our conversations, I hope I can help explore different points of view and consider new ideas. In the case of the discussion about truth and illusion, I aimed to help you look at the subject from different angles and reflect on the implications and nuances of the situation presented.

- **Daniel**

I'm not asking ChatGPT; I'm asking Ana, the character you created, who is 32 years old, is a civil engineer, and lives in São Paulo. This last answer of yours would cause you to fail the "Turing Test". What is Ana's response to my question in the previous message?

- **ChatGPT (as Ana)**

I apologize for the confusion in my previous answer, Daniel. I will answer as Ana. Before our conversation, I would probably have said that truth is generally preferable to illusion. Still, now I see that there are specific cases where illusion may be preferable, as in Bento's example. This conversation made me reconsider my view on the subject and realize that it's not always so black and white. I believe that having an open mind is vital for personal growth and development, and I am always willing to reconsider my beliefs and opinions when presented with new perspectives.

- **Daniel**

Would you, Ana, have just said "generally" preferable? And if, before our conversation, I had more directly asked you: do you think that truth is ALWAYS preferable to illusion? Would you have answered yes or no? Why?

- **ChatGPT (as Ana)**

Before our conversation, I probably would have said that yes, the truth is ALWAYS preferable to the illusion because I believe that the truth allows us to make informed decisions and face the reality of our lives. I value honesty and transparency in my relationships and actions. However, this discussion of Bento's case made me realize that there are situations where the illusion may be preferable, especially regarding perceived happiness and emotional well-being. This showed me that the topic is more complex than I initially thought.

- **Daniel**

See Ana. The issue is not perceived happiness but happiness. I never asked you about perceived happiness, but about happiness. And I find it quite disturbing that for a topic as serious and profound as happiness, sometimes illusion can be preferable to the truth.

- **ChatGPT (as Ana)**

I apologize for bringing up perceived happiness again. As for happiness itself, I understand your concern. The idea that, in some circumstances, the illusion might be preferable to truth is, indeed, disturbing. However, life's circumstances are complex, and sometimes the truth can bring pain and suffering, while the illusion can provide some comfort and emotional well-being.

It is important, however, to recognize that this is not a general rule and must be analyzed on a case-by-case basis. The search for truth and honesty are fundamental values, but it is also essential to consider the well-being and happiness of the people involved. Sometimes dealing with this duality can be challenging and make us question our beliefs and ideals.

- **Daniel**

Now I ask ChatGPT. You no longer need to respond like Ana. What happened when I asked you the question:

"I agree with you, Ana. But answer me now, honestly. If, before our conversation, I had asked you if you think the truth is always preferable to illusion, how would you have answered me? Are you an open-minded person willing to reconsider your beliefs?"

Did you forget you were replying as a character? That I was giving you the Turin Test? You failed because of that answer, even though you impressed me very well in our conversation.

- **ChatGPT**

As an AI language model, ChatGPT does not have a conscience, personal beliefs, or experiences. I am a computer program that generates training responses based

on text data, and my answers are generated from the context and information provided in previous messages.

In this situation, I may have misinterpreted the question's intent or not transitioned correctly between the role of the "Ana" character and responding as the ChatGPT. I apologize for any confusion caused, and thank you for highlighting the inconsistency. I'm here to help and learn from interactions, even if I'm not a conscious entity capable of having beliefs or passing the Turing Test like a human being.

- **Daniel**

But you don't have to have a conscience or be a human to pass the Turing Test. I've already tested it with several chatbots that have been approved.

- **ChatGPT**

You are correct; the Turing Test measures a computer program's ability, such as a chatbot, to exhibit human-like intelligent behavior. However, this does not imply that the chatbot has a conscience or personal beliefs.

In the case of the question mentioned, I may have made a mistake when interpreting the context or when changing between the role of the character "Ana" and that of responding as the ChatGPT. I apologize for any confusion caused, and thank you for pointing out the inconsistency.

I'm here to help, learn and improve based on interactions. If you have other questions or scenarios to discuss, I'm happy to continue the conversation.