

## A Summary And Response to Bernard Williams' Take on The Personal Identity Problem

Suppose someone greets you off the street as you walk to work, just as you do every morning. They approach you and ask, "Who are you?" You take a breath in preparation to respond, only to get told that they do not need a verbal answer, that it is a rhetorical question. You shrug and head on your way.

A week later, you are approached again by the same person at the same time and place and they ask if you remember who they are. You respond, "Yes I do, you asked me this time last week who I was."

They smile and nod to reassure that your answer was correct. They respond, "Yes, and this week I am going to ask you if you are the same person that I asked a question to last week."

You say, "Well, obviously I am the same person that you approached last week."

"What makes you the same person?" They respond.

"Well, what do you mean?" You are curious as to what they just asked. You are the same person, they recognized your physical appearance to approach you again knowing you two had spoken before, and you had recollection in your mind of being asked a question last week ensuring that the person they once spoke to was in fact you.

"I mean, what makes the person I am speaking to now the same person I briefly spoke with last week? Who are you?" And with this response, you understand.

One may reflect on this situation in one of three ways:

- I. They may agree with Descartes and say that their thinking mind, one that is closely conjoined with the body but could live without it, is what makes you *you*. They may take this further into the idea of self-consciousness and memories. The

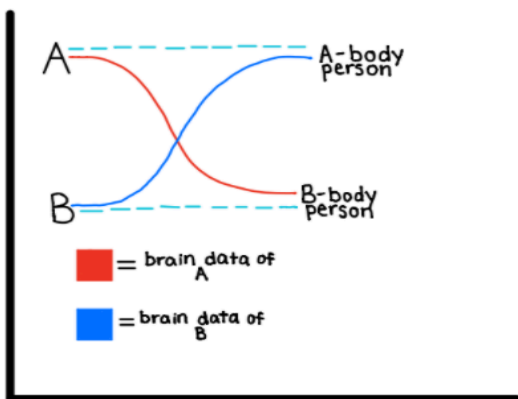
memory of yesterday was experienced in the way you remember it, indicating that the **person who you are remembering** is one and the same as the **person who is doing the memory recall**.

- II. They may conclude that what makes them the same person is their body, the skeleton, muscles and skin that express their mental states in a unique way. The person that walked yesterday is the same person walking today in the exact same stride if no other complications occur.
- III. Or, they may mesh 1 and 2 together to claim that identity cannot alone be thought of as psyche or physical, but rather a combination of both. Your mental states connect with your body to perform unique actions that are inspired by memories and individualized philosophies or worldviews.

Bernard Williams proposes a thought experiment that is designed to illustrate a way to see the arguments behind the answers written above, and in the end conclude that what makes you the same person is your body, and agree with the bodily theory. I find myself disagreeing and instead argue that the indications of self identity are not the physical extensions of our mind but rather the mind itself.

Williams begins by asking the reader to imagine a scenario in which a person labeled A “exchanges bodies” with a person labeled B (Williams 1). The memories and mental states of A are now in B, while the memories and mental states of B are now in A. There are complications to this that Williams addresses as we move forward, such as intention and performance. B’s character now performing actions in A’s body might be impossible if they are not alike. If gender, body type, etc. are too different for the original B mannerisms to be performed, then they may be unrecognizable. Williams asks us to forget this and instead say that A and B are as alike

as it is necessary for this thought experiment to continue. So, Williams says that if technology were constructed in a way to download a brain's data in some machine and swap it with the data of another into two bodies, we now have two people with exchanged bodies. A and B walked into this machine, and now B's brain data is in A's body and is identified as A-body-person. A's brain data is now in B's body to be identified as A-body-person. Visually, this can be represented in Figure 1:



**Figure 1**

This figure shows that now A has the body that was initially B's. That is, B-body person consists of B's original body and A's brain data. The dotted line is used to represent the physical body of A and B and how the bodies traveled on the graph. We will see, through William's argument, why he thinks that the dotted line is what indicates personal identity (the definition of what makes you the same as you over time).

Now, Williams addresses intention: why would two people undergo this process, and what do they get out of it? This is a necessary point and ends up being the *main* point for his argument. The scientist running this research project tells both A and B that after the process of exchanging bodies is over, one person will receive 100,000 dollars and the other will be tortured. He asks both participants who they would prefer to receive the money, and who they would prefer to receive the torture. This is designed to illustrate that the intuitive answer would be that

whoever the participant identifies themselves to be after the experiment is the participant they would prefer to receive the money. Williams calls the money a pleasant outcome or treatment, and calls the torture an unpleasant outcome or treatment. Let's consider A saying, "I would like the A-body person to receive the unpleasant outcome, and B-body person to receive the pleasant outcome," while B says, "I would like the A-body person to receive the pleasant outcome and the B-body person to receive the unpleasant outcome." Now, it is impossible for both A and B to receive the outcome they desire. For only one can receive the money and the other will receive torture. If A-body person were to receive the money and B-body person were to receive torture, then it would *seem* that A would not be getting what they desired and would say so, as they have their memories in the B-body person and would remember requesting the reverse outcome. A would say that they did not want the torture happening to *them*, indicating that the psyche is what identifies them. Williams agrees with this in saying that it would appear that after the experiment, B-body person is A, and A-body person is B. This would then indicate that A chose the best outcome for himself, to give B-body person the money, and has the right to be upset that he did not receive that outcome.

Now, let us focus on A, which up to this point seems to be B-body person. We can consider this as such since up until the present Williams has presented a thought experiment with intuition pointing towards A seeming to be B-body person. We will see that this quickly shifts, and Williams instead attempts to prove that A is A-body person, just as the dotted line represents in figure 1. He does so by saying that not only will A undergo torture, but further steps will be taken prior to the torture. The point of this, which I will highlight before saying the steps rather than after, is to question: at which point in the steps prior to torture would one intuitively say that this is a different person? Williams shows that at none of the steps would you say that it is a

different person, concluding that what makes a person the same person over space and time is their body. Relating to the experiment, this would mean that before and after the experiment, A is the same as A-body person. Now let us review the steps. I shall quote them exactly as Williams has written them:

“

- (i) A is subjected to an operation which produces total amnesia;
- (ii) amnesia is produced in A, and other interference leads to certain changes in his character;
- (iii) changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
- (iv) the same as (iii), except that both the character traits and the "memory" impressions are designed to be appropriate to another actual person, B;
- (v) the same as (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;
- (vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction.

“ (13). So, A is told that they will forget they are being tortured prior to it occurring. This fact does nothing to alleviate A's fear; for the situation is still going to occur despite remembering it right before or not. A is further told that forgetting the torture will be accompanied by forgetting things he presently (present meaning as he is being told these steps) knows. This also does not alleviate any fear. Being told that the memories they will have will not be ones they can recognize presently, but instead will be identical to memories of another person known as B,

does not alleviate any fear either. What does this prove about personal identity? Well, the idea is that as the participant is being told all of the steps, the fear is never alleviated. The participant is feeling this fear because they are aware that the torture will be happening to *them*. No psychological changes makes the participant think, “Ok, so the torture will happen to someone else.” Instead, the participant is consistently fearful of the torture that will happen to the same person who is listening to the description of events. If this is the case, then what makes the participant the same person over time is their body (according to Williams).

Williams is asking this: Does one get confused when the steps are being listed and the researcher uses the word *you* to communicate with the participant? It surely seems not. Does the participant consistently feel fearful in the six steps due to the dooming fact that *they* will receive torture? It seems surely so. Does the participant still feel the fear throughout all six steps because they believe that the torture will be happening to *them*? It seems surely so. So, in conclusion, wouldn't that support the bodily theory; wouldn't the body be how one is the same over time and space, if the participant undergoes numerous psychological changes and yet still intuitively feels that *they* will be the ones experiencing torture? It seems surely so.

This is to support that in figure 1, despite any changes that may alter the shape of the dotted line, a physical body is the determinant of someone being the same person as they were last week, or at any point in time and space.

I disagree with this and find myself believing that as drawn in figure 1, A is not in fact A body person but rather B body person. I want to begin by discussing a movie in which a thought experiment similar to Williams was executed. Black Box, a 2020 pop culture movie, involves a father named Nolan who got into a car accident. Nolan suffered from amnesia and his wife died after the accident. He had a lot of difficulty performing what used to be his everyday tasks, and

he found himself feeling as though his mantra and character are much different than how his friends and family described who he was before the accident. He decided to join a research study where the neurologist introduces a black box. The black box was designed to perform a mental exercise similar to hypnosis. This would allow for Nolan to recover his subconscious and access his mind. He attended multiple sessions and each time he would leave with more questions. Eventually, Nolan realizes that his subconscious is actually that of Thomas, who was the neurologist's son that had died. The doctor had downloaded Thomas' consciousness and implanted it into Nolan's body through a black box machine and hypnosis, which sounds awfully familiar to William's thought experiment. One of the larger differences is that Nolan was unaware that this had happened. Unlike the thought experiment explained by Williams, where both participants were given full disclosure about what would be happening, Nolan nor Thomas were aware that the consciousness of Thomas would be downloaded into the body of Nolan. However, the main push for Nolan's desire to participate in the study was that he felt so unlike himself- despite the body being his own and the outside world recognizing him, there was something about how he was processing the world that did not feel identical to the way he was processing the world before. And while amnesia may be a good reason for this, it went beyond the loss of memories- it was the overall perception that created new memories. I use this movie to show that despite being told Nolan was Nolan, and despite being supported in recovery by those who knew and loved him, Nolan knew he was not Nolan. He saw pictures of himself and the fact that the memory of the photo being taken was gone (due to amnesia) was not the complication at hand. Rather, it was the fact that Nolan did not feel like the Nolan in the photo. He had the same body as Nolan, and yet *he was not the same person as Nolan*. The conclusion of the movie is heartwarming as the daughter's voice allows Nolan's consciousness to override

Thomas'. Nolan reappears and once again feels like himself. Does this not intuitively prove that what made Nolan post-accident the same person as Nolan pre-accident was his consciousness? It seems as though intuition would agree that Nolan's body is clearly not what made him the same person over time. The memories being gone and the unfamiliarity of his own home and space did not cause Nolan to be in a state of confusion. Rather, it was the way his loved ones described how they knew him to be. They would explain how he was a loving, caring husband that would do anything for his family. They would tell him how his eye for photography was unique and astounding. However, nothing about what they would say would resonate with his current consciousness. His present (present meaning during the state of confusion) experience of the world did not align or identify with the experience of the world he perceived before.

This is why I argue that a person's consciousness, their awareness and experience of the world, is the reason we stay the same people over time, and more specifically the perception of the world and internal states are what truly makes you the same person over time.

I understand that (as Williams says) there is an undeniability to a body's natural reaction despite countless different psychological changes. Natural reactions are those that we cannot control- i.e a bear jumping out at us and our body's natural response is to sweat, elevate the heart rate and breathing rate, release adrenaline, etc. But I argue that we are missing a large step before the body can even react. How does one know what torture is? Did they learn about it on TV, witness it first hand in a personal scenario, read about it in a book, learn about it in school? And when their brain processed the word, did it get stigmatized with a category universal to everyone else's? Surely not. Has this person experienced torture before? If they have, let us interview them and understand what it was like. But there is much fault to that! For one, we have



the issue of memory alone, which I will attempt to address. Furthermore, we have the inaccuracy of language and the inaccuracy of introspection.

### **Memory Issue**

A classic and timeless argument to the psychological view is the memory issue. The main reason being that memories are highly inaccurate, and this is very true. It has been proven that each time someone recalls a memory since it happened, the memory becomes less and less accurate each time it is recalled. This can be thought of like a piece of tape; if you stick it on a surface and keep picking it off and replacing it onto the surface, it will get less sticky each time. A memory may not be accurate, but let us ask ourselves the true importance of memory accuracy in terms of identifying oneself as the same person as yesterday. If someone were to ask you about your first day of school in fifth grade, the point is not about how accurate your story is. Rather, it seems more important to note that the person doing the recall is the same person who experienced that first day of school as only they can experience it. Well, one may respond that accuracy is important since someone may remember an experience of their first day of school and it in fact never happened, they were never there. Or they may say that it can be a big problem to associate oneself in a memory that did not happen since they are then defining themselves as the same person from an experience that does not exist. To this I say that they are missing the point. The definition is not based on the memory itself, but the experience of the memory, the perception of the memory, the way in which someone thinks of their life and the world. Sure, my accuracy of fifth grade's first day is off base, but can you experience it as I can? Surely not. I can experience my memory of fifth grade the same today and the same tomorrow. Even if I wake up with amnesia, the way I perceive the world to make new memories involves reactions only I know to experience. The makeup of who I am is not erased when my memories are. The way I

perceive a bluebird is different and unrecognizable as the person who is perceiving the bluebird with me. No matter what amnesia I go through, no matter what physical changes I may undergo, and no scale of accuracy on my recall of the bluebird changes how only I perceive that bluebird. And I will continue to experience the bluebirds, people, *life*, as only I do. Even if someone were to end up exactly like me, suppose they experience the world in an identical way, how could you prove that? In the introspection and language section, it seems impossible to prove that my experience is the same as another. I can never experience any other mind except for my own. We do not know enough about the brain to prove that mental processes are the same in any sort of scientific way. We could only prove this by word of mouth and that is not at all a reliable source. Until that changes, I am the same person as I was a week ago thanks to my personal experience.

### **Introspection and Language**

\_\_\_\_\_ I now want to discuss why no one will be able to experience/perceive life as another, or be able to explain their perception for accurate understanding. Introspection is the process of explaining what is going through your head as it is happening. It is describing your mental processes and feelings as they occur in the moment. This idea has been tested on both professionals (considered to be highly emotionally intelligent) and average people who do not consider emotions as often as professionals. The results showed that in either case, professional or not, humans are highly inaccurate when it comes to introspection. In other words, there is no reliability in describing our own mental state.

Let us say that somehow, there was someone who was accurate at introspection. This person is able to perfectly explain their experience and understands themselves enough to be very accurate at introspecting their brain. Even if this was the case, language prevents this person from allowing another to accurately understand what is in their brain. What I mean by this is that

there is no plausible way (with the amount of emotions we have and the varying degree of each emotional experience) to articulate a mental process in any sort of reliable way. So, even if introspection were possible (which it is not because humans cannot understand their brain processes enough), getting someone else to know your mind or thoughts is impossible. This is all to show that you are unique to yourself in your own brain, no one can experience life as you do or ever get to understand how your life is lived except for you. Well, what makes you, you? The one doing the thinking, your consciousness perceiving the world and talking to you as it does it.

One may argue that the way we experience things changes over time. For instance, an experience that I had at 5 years old would cause me to react differently than I would at 21 years old. However, this is the entire point; we can adapt our minds but we cannot change how they think. We can grow and evolve our thought processes but we cannot change innately how our consciousness decides to develop and become. I may change how I react to a situation due to developmental growth, and yet at both instances my unique perception of the world is what caused my reaction in both cases. Meaning, I was the same person as I was when I was 5 years old because even though I may have responded in x way and now I would respond in y way, the perception and mental process of coming to the conclusion of that response is the same. Sure, it may be evolved now, but it is not a different stream of consciousness. It is instead an evolved one.

\_\_\_\_\_I ask now for one to refer back to the opening scenario of this essay. And as you impact the world in only the way you can, ask yourself who you are. Ask yourself what makes you the same person as you were yesterday. Why do you tell a story from 10 years ago in the first person, insinuating that the person in the story is the same person telling it? Understanding this could

help unlock how we understand others, how we empathize in conflict, or how we can connect in a disconnected era.