

THE UNIVERSITY OF CALGARY

Ignorance is Bliss: The Case Against State of the World Theories of a
Good Life

by

Dario Cankovic

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

August, 2007

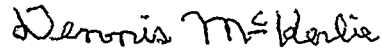
© Dario Cankovic 2007

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

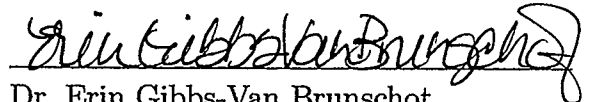
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Ignorance is Bliss: The Case Against State of the World Theories of a Good Life" submitted by Dario Cankovic in partial fulfillment of the requirements for the degree of MASTER OF ARTS.



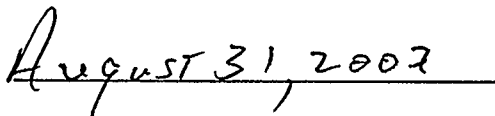
Supervisor, Dr. John A. Baker
Department of Philosophy



Dr. Dennis E. McKerlie
Department of Philosophy



Dr. Erin Gibbs-Van Brunschot
Department of Sociology



Date

Abstract

Theories of a good life can be partitioned into *state of the world* theories and *state of mind* theories. In recent times the uncritical acceptance of the conclusions Robert Nozick's *experience machine* thought experiment have resulted in the ascendance of *state of the world* theories over *state of mind* theories. In this thesis I proceed to explicate the debate between *state of the world* and *state of mind* theories, and analyze the *experience machine* thought experiment. I argue that the *experience machine* argument fails to establish the implausibility of *state of mind* theories, as well as that upon closer analysis *state of the world* theories are more problematic than their proponents admit. As such, we need to more seriously consider *state of mind* theories of a good life.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. John A. Baker, whose guidance over the course of both my undergraduate and graduate career has been invaluable, and without whose prodding this thesis would not have been finished in time. I would also like to express my gratitude to Dr. Dennis E. McKerlie, not only for agreeing to be on my thesis committee, but for also introducing me to philosophy; it was because of courses which I took with him in my first year that my love-affair with philosophy began. As well, I would like to thank Dr. Erin Gibbs-Van Brunschot for taking time out of her busy schedule to be on the thesis committee.

I am indebted to my parents, Elvira and Mladen Cankovic, for both their moral and financial support over the course of my education; they have financed, and continue to finance, my academic pursuits, even if at times they do not understand my reasons for pursuing something as, in their eyes, impractical as philosophy. In addition, I express my gratitude to my grandmother, Halima Atajic, who is a continuous source of inspiration to me.

Furthermore, I would like to thank Stephen Morgan for helping me edit the thesis, as well as for the tech-support he provided while I was writing the thesis. Also, my thanks go out my friends: Nico Beudeker, Alan Byers, John Chiu, John Lawless, and Joshua Shire, both for looking over various drafts of the thesis, as well as for the hours upon hours of philosophical conversation which they let me indulge in.

Table of Contents

Approval Page	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
1 Introduction	1
1.1 <i>State of Mind</i> vs. <i>State of the World</i>	2
2 The <i>State of the World</i> and a Good Life	12
2.1 General Form of the Argument	12
2.2 The Experience Machine	15
2.2.1 Desires	18
2.2.2 Reasons	21
2.3 Parallel Lives Thought Experiment	29
3 <i>Contra</i> State of the World Theories of a Good Life	34
3.1 Subjectivity and Objectivity	34
3.1.1 Problems for Objective State of the World Theories	37
3.2 The Experience Requirement	40
3.2.1 The Failure of the Experience Requirement	41
3.3 Implausibility of State of the World Theories	46
4 Defense of State of Mind Theories	49
4.1 States of Mind Always Matter	49
4.2 Why do we think States of the World matter?	51
4.2.1 States of the World Matter Most of the Time...	51
4.2.2 ...but NOT Always	54
4.3 The Agent's Own Evaluation of Their Life	57
4.4 The Simulation and Skepticism	59
4.4.1 Truth and a Good Life	62
5 Conclusion	65
Bibliography	68

Chapter 1

Introduction

Robert Nozick, in *Anarchy, State, and Utopia*, introduces a, now famous, thought experiment, that of the *experience machine*.¹ The purpose of the *experience machine* thought experiment is to refute, or minimally undermine, *state of mind* theories of a good life. Nozick and many other philosophers—such as James Griffin², David Brink³, Stephen Darwall⁴, and L.W. Sumner⁵—contend that, in the words of Sumner, “any theory with [the] implication [that whether or not an agent has a good life is entirely determined by features of the experience which are available to introspection—how it feels, how agreeable we find it, how much we wish it to continue, or whatever] is too interior and solipsistic to provide a descriptively adequate account of the nature of [the good life]”.⁶ That is, negatively, whether or not an agent has a good life does not consist *solely* in the mental states of that agent, and, positively, the *state of the world* plays a role in determining whether or not an agent has a good life.

In this thesis I will respond to the arguments of Nozick *et al.* I shall closely examine the argument employed by Nozick *et al.* against *state of mind* theories of a good life, and proceed to argue that they do not constitute a refutation of *state of mind* theories of a good life. Furthermore, apart from the negative argument against

¹See Robert Nozick, “The Experience Machine,” in *Anarchy, State, and Utopia*, (Oxford: Basil Blackwell, 1974):42-45.

²Griffin, 9-10.

³Brink, 223-224.

⁴“Self-Interest and Self-Concern,” *Social Philosophy and Policy* 14 (1997): 158-178; 162, 178.

⁵Sumner, 94-98.

⁶*Ibid.*, 98.

Nozick *et al.*, I shall further provide independent arguments in favor of *state of mind* theories, as well as independent arguments against *state of the world* theories of a good life.

1.1 *State of Mind vs. State of the World*

Before proceeding we must first explicate what exactly *state of mind* theories say, and what exactly *state of the world* theories say, in order to fully understand the nature of the dispute.

Firstly, both *state of the world* and *state of mind* theories are theories of a good life, that is to say, they are theories of what it is that makes an agent's life a good life for that agent. *State of mind* theories make whether or not an agent's life is a good life depends *solely* on that agent's mental states; nothing apart from the mental states of an agent, whose life is being evaluated, factor into consideration. *State of the world* theories, conversely, maintain that whether or not an agent has a good life is *not solely* dependent upon that agent's mental states.⁷ Since both mental states and states of the world are *states of affairs*, the general form of both *state of mind* and *state of the world* theories may be formulated as follows:⁸

T₁ For any agent A, there is a state of affairs S, such that *if S obtains then A has a good life.*

OR

⁷Ibid., 82.

⁸My thanks to my supervisor, John A. Baker, for the suggestion to formulate the general form of the arguments in the following fashion.

T₂ For any agent A, there is a state of affairs S, such that *if and only if* S obtains then A *has* a good life.

OR

T₃ There is a state of affairs S, for any agent A, such that *if* S obtains then A *has* a good life.

OR

T₄ There is a state of affairs S, for any agent A, such that *if and only if* S obtains then A *has* a good life.

OR

T₅ For any agent A, there is a state of affairs S, such that *if* S *does not* obtain then A *does not* have a good life.

OR

T₆ For any agent A there is a state of affairs S, such that *if and only if* S *does not* obtain then A *does not* have a good life.

OR

T₇ There is a state of affairs S, for any agent A, such that *if* S *does not* obtain then A *does not* have a good life.

OR

T_8 There is a state of affairs S , for any agent A , such that *if and only if* S does not obtain then A *does not* have a good life.

Which of the abovementioned forms, T_1 through T_8 , a theory of a good life takes will depend upon the specific theory. T_1 , T_2 , T_5 , and T_6 are agent-relative accounts, insofar as for any agent A the state of affairs S which makes that agent's life a good life need not make another agent's life good, the state of affairs is relative to the agent. That is, given two agents A and B , there will be some state of affairs S_A and some state of affairs S_B , such that S_A makes A 's life good (or the absence of S_A makes A 's life fail to be good), and likewise S_B makes B 's life good (or its absence makes B 's life fail to be good), where S_A and S_B are not equivalent, where $S_A \neq S_B$. For example, a *preferentist* theory of a good life, which makes whether or not an agent has a good life dependent upon whether or not the desires of said agent are satisfied, would be an agent-relative theory of a good life, and would fit one of the general schemas T_1 , T_2 , T_5 , or T_6 , insofar as the state of affairs which would have to obtain in order for an agent's life to be made good would be relative to each agent's specific desire set.

Conversely, T_3 , T_4 , T_7 , and T_8 are agent-neutral accounts, insofar as the state of affairs S which makes an agent's life good (or its absence makes an agent's life fail to be good) is the same for all agents. That is, given two agents A and B , the state of affairs S_A and the state of affairs S_B , where S_A makes A 's life good (or the absence of S_A makes A 's life fail to be good), and likewise where S_B makes B 's life good (or its absence makes B 's life fail to be good), will be equivalent, $S_A = S_B$. For example, a classical *hedonistic* account is an agent-neutral account of a good life, insofar as

it makes whether or not an agent has a good life dependent upon the pleasure in that agent's life, such that the state of affairs that would have to obtain in order for an agent to have a good life would be the same for all agents, namely, the state of affairs that would have to obtain in order for an agent to have a good life would be pleasure.

The above general formulation of theories of a good life does not, in it of itself, explicate the differences between *state of mind* and *state of the world* theories of a good life, insofar as both mental states and states of the world are states of affairs. As such, a more precise formulation of each theory would be as follows; in the case of *state of the world* theories:

W₁ For any agent A, there is a state of the world W, such that *if W obtains* then A *has* a good life.

OR

W₂ For any agent A, there is a state of the world W, such that *if and only if W obtains* then A *has* a good life.

OR

W₃ There is a state of the world W, for any agent A, such that *if W obtains* then A *has* a good life.

OR

W₄ There is a state of the world W, for any agent A, such that *if and only if W obtains* then A *has* a good life.

OR

W_5 For any agent A, there is a state of the world W, such that *if W does not* obtain then A *does not* have a good life.

OR

W_6 For any agent A there is a state of the world W, such that *if and only if* W *does not* obtain then A *does not* have a good life.

OR

W_7 There is a state of the world W, for any agent A, such that *if W does not* obtain then A *does not* have a good life.

OR

W_8 There is a state of the world W, for any agent A, such that *if and only if* W *does not* obtain then A *does not* have a good life.

One might attempt to formulate *state of mind* theories in the following fashion:

M_1 For any agent A, there is a mental state M of A, such that *if M obtains* then A *has* a good life.

OR

M_2 For any agent A, there is a mental state M of A, such that *if and only if* M *obtains* then A *has* a good life.

OR

M₃ There is a mental state M of A, for any agent A, such that *if M obtains* then A *has* a good life.

OR

M₄ There is a mental state M of A, for any agent A, such that *if and only if M obtains* then A *has* a good life.

OR

M₅ For any agent A, there is a mental state M of A, such that *if M does not obtain* then A *does not* have a good life.

OR

M₆ For any agent A there is a mental state M of A, such that *if and only if M does not obtain* then A *does not* have a good life.

OR

M₇ There is a mental state M of A, for any agent A, such that *if M does not obtain* then A *does not* have a good life.

OR

M₈ There is a mental state M of A, for any agent A, such that *if and only if M does not obtain* then A *does not* have a good life.

M_1 through M_8 fail to fully capture the claims of *state of mind* theories. M_1 , M_3 , M_5 , and M_7 are fully compatible with *state of the world* theories insofar as they do not preclude states of the world W from *also* influencing the life of an agent. *State of mind* theories are more properly formulated as negations of *state of the world* theories, insofar as beyond the claim that mental states affect the life of an agent, *state of mind* theories also claim that mental states *alone* determine whether or not an agent has a good life or not. As such, we must further add the following to the claims of *state of mind* theories:

$\neg W_1$ It is NOT the case that for any agent A , there is a state of the world W , such that *if W obtains then A has a good life.*

AND

$\neg W_2$ It is NOT the case for any agent A , there is a state of the world W , such that *if and only if W obtains then A has a good life.*

AND

$\neg W_3$ It is NOT the case there is a state of the world W , for any agent A , such that *if W obtains then A has a good life.*

AND

$\neg W_4$ It is NOT the case there is a state of the world W , for any agent A , such that *if and only if W obtains then A has a good life.*

AND

$\neg W_5$ It is NOT the case for any agent A, there is a state of the world W, such that
if W does not obtain then A does not have a good life.

AND

$\neg W_6$ It is NOT the case for any agent A there is a state of the world W, such that
if and only if W does not obtain then A does not have a good life.

AND

$\neg W_7$ It is NOT the case there is a state of the world W, for any agent A, such that
if W does not obtain then A does not have a good life.

AND

$\neg W_8$ It is NOT the case there is a state of the world W, for any agent A, such that
if and only if W does not obtain then A does not have a good life.

A *state of mind* theory can thus properly be construed as M_1 , or M_2 , or M_3 , etc. in conjunction with the negation of state of world theories, i.e. in conjunction with $\neg W_1$ through $\neg W_8$.

Since on any *state of mind* theory of a good life whether or not an agent has a good life is *solely* dependent upon an agent's mental states, any *state of mind* account must deny W_1 through W_8 . Whereas, *state of the world* theories need not deny that mental states play a part in determining whether or not an agent has a good life, insofar as they also allow that states of the world play a role as well. As such, *state of the world* theories are compatible with M_1 , M_3 , M_5 , and M_7 insofar as

none of these forms of *state of mind* accounts explicitly preclude states of the world from playing a part in determining whether or not an agent's life is good.

Having explicated what *state of mind* and *state of the world* theories state we can now move on to further explicate what the disagreement between the two sets of theories amounts to.

The disagreement between *state of the world* and *state of mind* theories can be formulated more precisely in the following fashion:⁹ for any agent A and for any agent B who live psychologically (i.e. cognitively, affectively, and conatively) identical lives, where P_A is the propositional content of some psychological (cognitive, affective, or conative) state of agent A, and P_B is the propositional content of some psychological (cognitive, affective, or conative) state of agent B, and the psychological states which P_A and P_B are propositional contents of are in some way relevant to the question of whether A and B are having good lives, let it be the case that $\neg P_A$ and further let it be the case that A believes that P_A , and also let it be the case that P_B and that B believes that P_B . On a *state of mind* theory the failure of the propositional content of some psychological state of an agent to obtain, i.e. that $\neg P_A$, has no impact on the life of the agent in question, i.e. A. Furthermore, still more strongly, that $\neg P_A$ not only *does not* have an impact on the life of A, but it *cannot* alone have an impact on A's life; where here alone means without A somehow coming to be aware of the fact that $\neg P_A$, that is, without A coming to believe that $\neg P_A$. In other words, on a *state of mind* theory the life of A and the life of B are and must be equally good.

Conversely, on a *state of the world* theory, *ceteris paribus* the life of A and the life of B *are not* and *cannot* be equally good. The life of A would necessarily be

⁹Once again, thanks to my supervisor, John A. Baker.

less good than that of B insofar as it is the case that $\neg P_A$. Even though A believes that P_A , and in virtue of this belief is psychologically identical to B, the fact that $\neg P_A$, where P_A is the propositional content of some psychological state relevant to the evaluation of A's life, has a detrimental impact on A's life even though A is not aware of it.

Chapter 2

The *State of the World* and a Good Life

2.1 General Form of the Argument

Those who argue for a *state of the world* theory of a good life do so primarily with negative arguments directed against *state of mind* theories of a good life. We might construct the general form of the argument in the following fashion:¹⁰

- 1 Let some *state of mind* theory M^* of the good life be true. [Assumption for *modus tollens tollendo*]
- 2 Then, given the statement of the *state of mind* theory, it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is *not* having a good life *if* P_i does *not* obtain. [from 1, given the characterization of the *state of mind* theory, i.e. given $\neg W_1$ through $\neg W_8$, specifically given $\neg W_5$]
- 3 Let persons A and B live psychologically (i.e. cognitively, affectively and conatively) identical lives. [Assumption]
- 4 Let P_A and P_B be states of the world such that:
 - (a) P_A is the propositional content of some cognitive, affective, or conative state of person A and P_B is the propositional content of some cognitive, affective, or conative state of person B, AND

¹⁰Once again, thanks to John A. Baker for the general form of the state of the world argument.

(b) the cognitive, affective, or conative states of which P_A and P_B are propositional contents be in some way relevant to the question of whether A and B are having good lives. [Assumption]

5 Let it be the case that $\neg P_A$ without A knowing this, but let it be the case that P_B .

6 Then, despite 5, both A and B are having equally good lives.

7 BUT, 6 is not plausible [see e.g., the experience machine]

8 Therefore, 2 is not plausible and hence by *modus tollens tollendo*, 1 is not plausible.

The formulation of premise 2 is derived from $\neg W_5$, though we may make the argument apply to any of W_1 through W_8 by substituting premise 2 with one of the following:

2A Then, given the statement of the state of mind theory, it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if* P_i obtains. [$\neg W_1$]

OR

2B Then, given the statement of the state of mind theory, it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if and only if* P_i obtains. [$\neg W_2$]

OR

2C Then, given the statement of the state of mind theory, it is *not the case* that there is a state of the world P_i such that for any agent A_i , A_i is having a good life *if* P_i obtains. $[\neg W_3]$

OR

2D Then, given the statement of the state of mind theory, it is *not the case* that there is a state of the world P_i such that for any agent A_i , A_i is having a good life *if and only if* P_i obtains. $[\neg W_4]$

OR

2E Then, given the statement of the state of mind theory, it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is *not* having a good life *if and only if* P_i does *not* obtain. $[\neg W_6]$

OR

2F Then, given the statement of the state of mind theory, it is *not the case* that there is a state of the world P_i such that for any agent A_i , A_i is *not* having a good life *if* P_i does *not* obtain. $[\neg W_7]$

OR

2G Then, given the statement of the state of mind theory, it is *not the case* that there is a state of the world P_i such that for any agent A_i , A_i is *not* having a good life *if and only if* P_i does *not* obtain. $[\neg W_8]$

2.2 The Experience Machine

The general argument given above depends upon premise 7, that is, it depends upon establishing the implausibility of *state of mind* theories of a good life, where establishing the implausibility would consist in showing that *state of mind* theories of a good life are incompatible with our common sense views about what makes an agent's life a good life for that agent. In order to establish this implausibility an independent argument must be provided. Such an argument is to be found in the form of Nozick's *experience machine* thought experiment; it attempts to establish the implausibility of *state of mind* theories of a good life.

Nozick asks us to:

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug yourself into this machine for life, preprogramming your life's experiences? [...] Of course, while you are in the tank you won't know that you're there; you'll think it's all actually happening. [...] Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few moments of distress compared to a lifetime of

bliss, and why feel any distress at all if your decision is the best one?¹¹

Nozick goes on:

What does matter to us in addition to our experiences? First, we want to *do* certain things, and not just have the experience of doing them. In the case of certain experiences, it is only because we first want to do the actions that we want the experiences of doing them or thinking we've done them. [...] A second reason for not plugging in is that we want to *be* a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, witty, loving? It's not merely difficult to tell; there's no way he is.¹²

We may reconstruct Nozick's argument from this paragraph in the following fashion:

- 1 Let some *state of mind* theory M^* of the good life be true. [Assumption for *modus tollens tollendo*]
- 2 Then, given the statement of the state of mind theory, the only thing that would *matter* to an agent is how their life *feels from the inside*, i.e. the only thing that could *matter* to an agent are mental states.

¹¹Nozick, 42-43.

¹²Ibid.

3 BUT, it is *not the case* that the only thing that *matters* to an agent is how their life *feels from the inside*, i.e. their mental states, since:

(a) agents want to *do* certain things, they do *not* merely want to have an *experience* of doing them; AND

(b) agents want to *be* a certain way, they do *not* merely want to have the *experience* of being a certain way.

4 Therefore, 2 is not true and hence by *modus tollens tollendo*, 1 is not true.

A number of issues emerge with this argument. Firstly, premise 2 is suspect, or, minimally, in need of elaboration. Nozick supposes that if it is the case that a *state of mind* theory M^* of a good life is true then this would make it the case that the only thing that would *matter* to an agent are mental states. ‘Matters to an agent’ here is, presumably, to be understood in the following fashion: for any agent A_i a state of affairs P_i matters to agent A_i if (and perhaps even if and only if) P_i is the propositional content of some cognitive, conative, or affective state of A_i . Nozick further supposes that if M^* is true that an agent A_i would have some reason R_i to hook themselves up to the experience machine. Still further, he supposes that the reason R_i would be an overriding reason when he says: “What’s a few moments of distress compared to a lifetime of bliss, and why feel any distress at all if your decision is the best one?”¹³ R_i is an overriding reason for A_i if and only if A_i has no other reason R_x such that R_x outweighs, or is of equal weight to, R_i , i.e. such that $\neg(R_x \geq R_i)$. I shall argue that premise 2 is mistaken, that even if a *state of mind* theory M^* of the good life is true it does not follow that: (a) the only thing

¹³Ibid.

which will matter to an agent are mental states, and (b) that an agent has a reason to hook themselves up to the *experience machine*. A theory of the good life, *state of mind* or *state of the world*, is a *descriptive* account of what it is for an agent to have a good life, it need not be *normative*, and as such is disconnected from how an agent *ought* to live their life, it merely describes what a good life for that agent consists in. As such it can be the case that M^* is true while remaining the case that states of affairs other than mental states matter to an agent, as well as that an agent has no reason to hook themselves up to the *experience machine*.

2.2.1 Desires

Nozick is correct on one very important point, namely, that agents desire (where desires are synecdochically construed, and even when narrowly construed) more than just mental states. This is to say that the psychological states of some agent A_i , relevant to the evaluation of that agent's life, have as their propositional content states of affairs P_i beyond some mere mental states M_i . Let us, for sake of brevity, render a state of affairs P_i which is the propositional content of an agent's psychological (cognitive, conative, and affective) states, and which is relevant to the evaluation of that agent's life, as $A(P_i)$. Trivially $M_i \subset P_i$, that is, the set of all mental states M_i is a *proper subset* of all states of affairs P_i (unless of course one is an idealist and believes all states of affairs to be mental states, but we can safely disregard that position). Furthermore, it might be the case that $A(M_i) \subset A(P_i)$, that is, that all mental states M_i which are the propositional content of an agent's psychological states (relevant to the evaluation of that agent's life) are a *proper subset* of possible states of affairs P_i which are propositional contents of an agents psychological (cognitive, conative, and

affective) states; this would imply that an agent desires more actual states of affairs as opposed to mental states of affairs. Nevertheless this is not required for Nozick's claim, and in fact would be far too strong for Nozick. All that is required is that the set of $A(P_i)$ is not equivalent to the set of $A(M_i)$, that is, that the set of $A(M_i) \neq$ set of $A(P_i)$. That there could be some element P_i such that $P_i \in A(P_i)$, and $P_i \notin A(M_i)$. While an agent might desire mental states as part of their subjective motivational set, $A(M_i)$, they also desire non-mental states, $A(P_i) \notin M_i$. This seems obvious, insofar as, for example, if I want to be loved presumably I want to *actually* be loved, as opposed to merely *believe* that I am loved. The *belief*, a mental state, is *not* the propositional content of my desire, rather the propositional content of my desire is the state of affairs of *being loved*. In short, the objects of my desire in many cases consist of states of affairs beyond mental states.

Nozick concludes from this fact, that the propositional content P_i an agent A_i 's psychological states is directed towards states of affairs apart from mental states M_i , that any *state of mind* theory M^* of the good life must be mistaken, or, minimally, is implausible. In order for this conclusion to follow, from the fact that the propositional content of an agent's psychological states is directed toward more than mere mental states, it would have to be the case that any *state of mind* theory M^* states that, if M^* were true, then the propositional content of an agent's psychological states would be directed *solely* toward mental states. That is, for Nozick's claim to follow, it would have to be the case that any *state of mind* theory M^* , in addition to claiming M_1 or M_2 or M_3 or M_4 , etc. AND $\neg W_1$ through $\neg W_8$, it would further have to claim that if M^* is true then for any agent A_i the propositional content of A_i 's psychological (cognitive, conative, and affective) states will be directed toward states of affairs P_i

which are mental states M_i , such that $A(P_i) = M_i$. Nevertheless, not only does any *state of mind* theory M^* fail to claim this, no *state of mind* theory M^* need make this assertion.

All M^* claims is that the mental states are the only things which affect whether or not an agent is living a good life or not, it makes no additional assertions regarding the propositional content of any agent's psychological states. An agent could still very well aim for and desire states of affairs P_i which are not mental states M_i on any state of mind theory M^* ; M^* would only have to maintain that whether or not P_i obtains, where $A(P_i) \neq M_i$, is entirely irrelevant to whether or not agent A_i has a good life. Thus, on a *state of mind* theory M^* any agent A_i might still be motivated to pursue states of affairs P_i which are not mental states M_i ; that is, states of affairs P_i which are not mental states M_i might still *matter* to A_i in the sense that A_i *cares* about whether or not P_i obtains; that is, in the sense that P_i is the propositional content of some of A_i 's cognitive, conative, or affective states. It would just be the case, if M^* were true, that whether or not P_i obtains has no impact on whether or not A_i has a good life or not, and, as such, would be irrelevant to the evaluation of A_i 's life.

This might seem *prima facie* implausible, that it could be the case that an agent *cares* about something, that is, they desire it, want it, enjoy it, etc. BUT that the something which the agent cares about can fail to have an impact on whether or not an agent is living a good life. Nevertheless, a mundane example might suffice to illustrate how it could be the case that what an agent *cares* about can be irrelevant to whether or not that agent has a good life. Suppose we have an agent who is deeply politically committed and cares about the outcome of the next American

presidential election. Before the presidential election the agent moves away to some tropical paradise in the South Pacific which has no access to the outside world; the agent is perfectly happy in their new home, enjoying the sunny beaches, swimming in the warm ocean, and just in general having fun and take pleasure in the simple things in life. Unknown to the agent the candidate that they wanted to win fails to win the election, and the candidate that they did not want to win succeeds in securing the presidency. Of course, the agent remains entirely *blissfully* unaware of this fact, and goes on enjoying life on the tropical paradise. In such a case we would be inclined to say that while the agent cared about the outcome of the presidential election, that the outcome was irrelevant to the agent's life; their life seems to have not at all been impacted by the outcome of the election. Now, of course, the example above is not nearly as radical as the position of any *state of mind* theory M^* which disregards any propositional content P_i which is not a mental state M_i in evaluating an agent's life, nevertheless it serves to illustrate the point that we do at times think that what an agent cares about is not relevant to the evaluation of that agent's life.

2.2.2 Reasons

After presenting us with the experience machine Nozick asks the poignant question "Would you plug in?" The question, while a very interesting one, is actually irrelevant to the consideration of what a good life consists in. Largely because it conflates what an agent has *reason* to do or refrain from doing, and what would make an agent well off; the two are not equivalent.

We have already granted Nozick that an agent A_i can have cognitive, conative, and affective states with some propositional content P_i which is not a mental state

M₄. That is, we have granted to Nozick that an agent's subjective motivational set contains desires for states of affairs beyond mental states.

Since:

Sentences of the forms 'A has a reason to Φ ' or 'There is a reason for A to Φ ' (where ' Φ ' stands in for some verb of action) seem on the face of it to have two different sorts of interpretation. On the first [internal reasons interpretation], the truth of the sentence implies, very roughly, that A has some motive which will be served or furthered by his Φ -ing, and if this turns out not to be so the sentence is false: there is a condition relating to the agent's aims, and if this is not satisfied it is not true to say, on this interpretation, that he has a reason to Φ . On the second [external reasons] interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive.¹⁴

In order to make the above Williams passage more intelligible and relevant in the context of this thesis we can reformulate what Williams is saying in the following fashion:

Sentences of the forms 'A has a reason to Φ ' or 'There is a reason for A to Φ ' (where ' Φ ' stands in for some verb of action) seem on the face of it to have two different sorts of interpretation. On the first [internal reasons

¹⁴Williams, 101; Williams himself talks first about reasons *interpreted internally* and reasons *interpreted externally*, and only latter chooses to refer to reasons *interpreted internally* as 'internal reasons', and likewise reasons *interpreted externally* as 'external reasons' for the sake of brevity, I shall follow the same convention here.

interpretation], the truth of the sentences implies that A has some state of affairs P_A as the propositional content of their psychological (cognitive, conative, and affective) states such that P_A will be further served by A's Φ -ing. On the second [external reasons] interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate state of affairs P_A which is the propositional content of A's psychological states which would further be served by A's Φ -ing.¹⁵

It follows from the fact that an agent A_i has an element D_i in their subjective motivational set such that the propositional content of D_i is a state of affairs P_i that A_i will be *motivated* to pursue P_i , and, given Williams, A_i will have some *internal* reason IR_i to pursue P_i .¹⁶ Failing that, that is, if it is not the case that an agent A_i has an element D_i in their subjective motivational set such that the propositional content of D_i is a state of affairs P_i , then A_i will *not* be motivated to pursue P_i , and hence would lack an *internal* reason IR_i to pursue P_i .

In the case of being presented with the option of hooking oneself up to the experience machine, whether or not one will do so is largely dependent on what reason one has for doing so. Ignoring for a moment the debate between internalism vs. externalism with regards to reasons, even if the fact (presuming for a moment that it is a fact) that I would be made extremely well off by hooking myself up to the experience machine were an external reason for me to do so, it remains the case that external reasons are motivationally inert unless we also have an equivalent internal reason.¹⁷ We can grant to Nozick that for some agent A_i , if hooking themselves up

¹⁵Ibid.

¹⁶Ibid.

¹⁷See Williams, "Internal and external reasons" in *Moral Luck* (Cambridge: Cambridge Univer-

to the experience machine would make their lives better off, as is contended by the *state of mind* theory, then this fact would constitute an *external* reason ER_i for A_i to hook themselves up to the machine. Nevertheless, it can still remain the case that A_i will fail to hook themselves up insofar as they also further lack a relevant *internal* reason IR_i to hook themselves up. In order for A_i to have an internal reason IR_i to hook themselves up to the experience machine it would have to be the case that A_i had some relevant motivation N_i ; where A_i has some relevant motivation N_i if and only if there is some element D_i in A_i 's subjective motivational set such that the propositional content of D_i is some state of affairs P_i which would obtain were A_i to hook themselves up to the experience machine. That is, in simpler terms, A_i would lack an IR_i to hook themselves up to the experience machine unless A_i has a desire for some state of affairs P_i which would *actually* obtain by connecting to the experience machine; since P_i is a desire for a state of affairs to obtain it would be insufficient if the experience machine merely provided the *feeling* of P_i obtaining. For example, if A_i were an agent who was entirely indifferent to whether or not their experiences were veridical or illusory, A_i 's subjective motivational set consisted solely of desires directed toward *mere* experiences, that is, A_i desired nothing besides *experiences*, then A_i would have an internal reason IR_i to hook themselves up to the experience machine insofar as the experience machine would *actually* satisfy A_i 's desires, that is, insofar as the propositional content of A_i 's desires would consist solely in states of affairs P_i which were mental states M_i , where $P_i = M_i$. In other words, the relevant motivation N_i of some agent A_i would have to have as its propositional content some mental state M_i in order for the experience machine to actually satisfy

sity Press, 1980).

N_i , and in order for A_i to have an internal reason IR_i to hook themselves up to the experience machine. Thus far this is trivial, insofar as all that has been said is that if the propositional content of an agent's psychological states is directed toward mental states then hooking that agent up to the experience machine would satisfy the propositional content of their psychological states. Nozick would not contest this, for his claim is that the agent A_i as described above would not have psychological states whose propositional content was directed *solely* toward mental states, but would instead include non-mental propositional content. That is, that there is no agent A_i such that all of the elements D_i of A_i 's subjective motivational set have as their propositional content states of affairs P_i which are mental states M_i .

Supposing that an agent A_i has psychological states whose propositional content is directed both toward mental states of affairs M_i and non-mental states of affairs P_i , where $P_i \notin M_i$, it is clear that an agent might have *some* reason, *internal* reason IR_i , to hook themselves up to the experience machine. Nevertheless, even if we grant that an agent A_i could have an internal reason IR_i to hook themselves up to the experience machine, that is, A_i *wants* to do so, other internal reasons IR_{x1} through IR_{xn} which A_i has might outweigh the reason IR_i to hook up. That is, even if there is an internal reason IR_i it does not follow that IR_i outweighs all other internal reasons IR_{x1} through IR_{xn} ; that is, it can still fail to be the case that $(IR_i > IR_{x1}$ through $IR_{xn})$, i.e. $\neg(IR_i > IR_{x1} \text{ through } IR_{xn})$. An agent A_i 's other reasons IR_{x1} through IR_{xn} could outweigh IR_i (such that IR_{x1} through $IR_{xn} > IR_i$). Nowhere does any *state of mind* theory M^* state that an agent will or must have internal reason IR_i to hook themselves up to the experience machine which outweighs all other internal reasons IR_{x1} through IR_{xn} . As such, it does not follow from the fact

that an agent lacks such an overriding reasons IR_i that M^* is false. For example, if an agent values *truth* and *authenticity* then that agent has reason to refrain from hooking themselves up to the machine, given that they would know *now* that the machine will provide them with *false* and *inauthentic* experiences. The fact that *after* the agent has hooked themselves up to the experience machine they will not be aware that these experiences are not *false* and *inauthentic*, that is, that they'll believe them to be *true* and *authentic*, does *not* negative that agent's reason *now* to refrain from hooking themselves up. The fact that an agent *would not* hook themselves up to the experience machine can therefore be accounted for by appeal to *internal* reasons of that agent which are entirely independent of whether or not that agent would be made better off by hooking themselves up to the machine. Hence we can concede to Nozick, and those who agree with him, that we might have (internal) *reasons* to refrain from hooking up to the experience machine without conceding that this implies that we would not be *better off* if we were to hook ourselves up.

To make this point clear consider the following example. Suppose you were offered the opportunity to go off to some far off country that you've always wanted to go to, and have also been offered a dream job there. Accepting this offer will make you considerably better off than you are now, but in order to accept the offer you must leave your family and friends and everything that you hold dear here and now behind. You might decline the offer even though it will make you better off, knowing fully that it will make you better off, because you value your friends, family, and all else you hold dear here and now more than you value your life going well for you. That is a simple example of where reasons for action, internal reasons, diverge from what would make your life go best for you. As such, the fact that an agent

would not hook themselves up to the experience machine does not undermine *state of mind* theories of a good life, since what you have reason to do, where reason is interpreted internally, is not equivalent to what would make your life a good life.¹⁸ As David Hume puts it: “‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. [...] ‘Tis as little contrary to reason to prefer even my own acknowledg’d lesser good to my greater, and have a more ardent affection for the former than the latter.”¹⁹

There is, thus, a gap between an agent’s *reasons for action* (i.e. *internal* reasons) and what would make that agent’s life better. While the fact that the agent’s life would be made better off might constitute an *external* reason ER_i (presuming for a moment that there are such things as external reasons) for that agent A_i to hook themselves up to the machine, unless that *external* reason ER_i has some corresponding *internal* reason IR_i , that is, unless the agent also *wants* to hook themselves up, the *external* reason ER_i will be entirely motivationally inert. Moreover, even if the agent has an *internal* reason IR_i to hook themselves up, it can still be the case that other *internal* reasons IR_{x1} through IR_{xn} outweigh or overpower that reason, i.e. (IR_{x1} through $IR_{xn} > IR_i$), hence resulting in the agent refraining from attaching themselves to the machine, and thus failing to make their life better. Thus, the question of whether you *should* plug yourself into the experience machine requires further elaboration. If the ‘*should*’ here is interpreted as ‘do you have reason to’, the answer depends upon whether we are talking about internal or external reasons. And in either case it can still be the case that you *should not* plug yourself into the

¹⁸Jason Kawall, “The Experience Machine and Mental State Theories of Well-being,” *The Journal of Value Inquiry* 33 (1999): 381-387; 382-83.

¹⁹David Hume, *A Treatise of Human Nature*, Book II, part III, section iii.

machine even if it is the case that you would be made better off by doing so. Hence, it does not follow that simply because you *should not* plug yourself into the machine that you would be made worse off by doing so. Thus, it is not the case that a *state of mind* theory M^* of a good life that an agent must either ‘always prefer’ the states of mind, over some state of the world, which will make that agents life the best off, nor is it the case that the agent ‘must be indifferent’ between real experiences and simulacra as long as they produce indistinguishable mental states in the agent.²⁰ An agent’s reasons and what would make that agent’s life best off are not equivalent, those who insist that they are, and thus that an agent *should* plug themselves into the experience machine, regardless of the agent’s actual preferences, are conflating the two.

The uneasiness and hesitation which Nozick speaks of when we are asked to contemplate attaching ourselves to the experience machine thus needs not be construed a product of an intuition which tells us that hooking ourselves up to the experience machine would not make us better off, but might instead be merely a result of conflicting reasons for action. As such, the experience machine thought experiment to which Nozick appeals is undermined insofar as we have an alternative explanation for what underlies an agent’s reluctance to hook themselves up to the experience machine, namely, the nonequivalence between what makes an agent’s life better off, and an agent’s reasons for action.

Therefore Nozick is mistaken in asserting that if some *state of mind* theory M^* of the good life is true then the only thing that matters to an agent are mental states, and that an agent would have reason, possibly overriding reason, to hook themselves

²⁰Sumner, 98.

up to the experience machine. As has been show it can be the case that M^* is true and that states of affairs other than mental states matter to an agent, as well as that an agent could lack reason to hook themselves up to the experience machine. Thus, since premise 2 of Nozick's argument is false the conclusion fails to follow. As such Nozick's experience machine thought experiment, at least the argument derived from it as presented here, fails to establish the implausibility of *state of mind* theories of a good life, and as such cannot serve as premise 7 of the *state of the world* proponents' argument against state of mind theories of a good life.

2.3 Parallel Lives Thought Experiment

Given the failure of Nozick's experience machine thought experiment to demonstrate the falsity, or minimally the implausibility, of *state of mind* theories of a good life, there is another argument to which a proponent of a *state of the world* theory of a good life can resort to. Instead of using Nozick's experience machine thought experiment to establish premise 7, i.e. the implausibility of *state of mind* theories of a good life, a proponent of *state of the world* theories can instead resort to a different thought experiment. Rather than considering a single agent in isolation and evaluating the life of that agent inside the experience machine, let us instead consider two agents in parallel, one inside the machine the other in the real world, i.e. one undergoing veridical experiences, and the other illusory; I shall refer to this as the *parallel lives* thought experiment, it is often employed in a variety of forms in debates concerning *state of mind* vs. *state of the world* theories of a good life.

In order to establish the implausibility, or falsity, of *state of mind* theories of a

good life, that is, in order to establish premise 7, this sort of thought experiment asks us to directly consider the lives of two agents, A and B, who are psychologically (cognitively, conatively, and affectively) identical. If the lives of the two agents are psychologically identical, then they will likewise be psychologically indiscernable, by Leibnitz's Law (indiscernibility of identicals, i.e. for any x and y, if x is identical to y, then x and y have all the same properties, rendered as $\forall x \forall y [x=y \rightarrow \forall P (Px \Leftrightarrow Py)]$). That is to say that internally the lives of A and B will be indistinguishable. A lives a life *outside* of the experience machine, whereas B lives a life *inside* the experience machine; thus, we may say, that A's experiences are *veridical* whereas B's experiences are *illusory*, insofar as B's experiences are not experiences of the real worlds, but are instead sufficiently advanced simulacra which are indistinguishable from experiences of the real world. We may well consider this sort of thought experiment a variant of Nozick's *experience machine*; in the experience machine thought experiment we were asked to consider only whether we would connect ourselves to the experience machine, that is, we were asked to consider whether or not the life of a single agent would be a good life were that agent connected to the experience machine; conversely, in the *parallel lives* thought experiment we are asked now to directly consider and contrast the lives of two agents, one inside the machine (i.e. undergoing illusory experiences), and one outside the machine (i.e. undergoing veridical experiences).

Consider some state of affairs P_A which is the propositional content of some cognitive, conative, or affective state of agent A, such that P_A is not a mental state M_A , i.e. $A(P_A) \neq M_A$. Likewise, consider some state of affairs P_B which is the propositional content of some cognitive, conative, or affective state of agent B, such that P_B is not a mental state M_B , i.e. $A(P_B) \neq M_B$. Furthermore, since agent A

and B are psychologically identical, let P_A be identical with P_B , that is, let $P_A = P_B$. Lastly, for the sake of argument, let P_A obtain, such that it is the case that P_A , where A believes that P_A obtains, i.e. where A's experience of P_A obtaining is veridical, AND let P_B fail to obtain, such that it is the case that $\neg P_B$, where B believes that P_B obtains, i.e. where B's experience of P_B obtaining is illusory. The *parallel lives* thought experiment now asks us to consider the parallel lives of agents A and B given the abovementioned conditions which have been outlined: *ceteris paribus*, do agents A and B live lives which are equally good, or is it the case that A's life is better than B's, or B's better than A's? This is the question that the *parallel lives* thought experiment asks us to consider.

The *state of the world* supporter who brings up the *parallel lives* thought experiment naturally wants us to answer that clearly A's life is, and cannot but be, better than B's given that the states of affairs P_A which is the propositional content of A's cognitive, conative, or affective states actually obtains, where the state of affairs P_B fails to obtain, such that it is the case that $\neg P_B$. Given that A and B are psychologically identical, such that any mental state M_A is identical to any mental state M_B , i.e. $M_A = M_B$, it cannot be the case that mental states of agent A and B would account for the difference between the lives of A and B, and therefore, it must be the case that the state of the world P_A obtaining while the state of world $\neg P_B$ obtains is what accounts for the difference between the lives. Hence, *state of mind* theories of a good life are false since they cannot account for A's life being better than that of B. Furthermore, if some *state of mind* theory M^* of a good life were true then it would be the case that A and B live equally good lives, but since that is, allegedly, not the case, any *state of mind* theory M^* of a good life must be false.

This is the general line of argument by *state of the world* proponents who employ the *parallel lives* thought experiment.

To make the argument more intelligible consider the following parallel cases. In the first case we are dealing with a parent (agent A) on their deathbed, whose child has gone off to war. An officer is sent by the military to inform the parent of their child's safe return. The officer, upon arrival, discharges their duty and informs the dying parent of the good news. The parent dies content knowing that their child is safe.

In the second case, everything is identical to the first case, except that the officer is charged with informing the dying parent (agent B) of their child's demise. Upon arriving he takes pity upon the parent, knowing that the parent themselves is going to die soon the officer decides to conceal the information of the child's death from the parent. Instead of reporting to the parent of their child's demise, the officer instead lies and tells the parent that their child is alive and well. The parent dies content *believing* that their child is safe.

The parent-child parallel case above is relevantly similar to the general form of the *parallel lives* thought experiment, insofar as we are dealing with two agents, A and B, who are psychologically identical, both of who desire that some state of affairs P_A and P_B obtain, where it is the case that P_A obtains with A being aware of P_A obtaining, and P_B fails to obtain, that is, $\neg P_B$ obtains, with B being unaware of $\neg P_B$ obtaining. The experiences of A are *veridical*, while the experiences of B are *illusory*.

Most people would be inclined to side with the *state of the world* proponent in evaluating the life of agent A, the parent whose child survives, as having gone better

for A than the life of agent B, the parent whose child dies. They reason that when we desire something we desire that the something in question *actually* obtain, as opposed to merely *believing* that it obtains, or having an *experience* of it obtaining. That is to say that the propositional content of our psychological (cognitive, conative, and affective) states is directed toward non-mental states of affairs $P_i \notin M_i$ (alongside mental states of affairs M_i). There is, thus, a gap between the *object of desire*, P_i , and a mere *experience* of the object of desire obtaining when it *actually* fails to obtain, that is, a mental state M_i such that the agent A_i believes that P_i when in fact it is the case that $\neg P_i$. In virtue of this most people would evaluate the life of agent A as having gone better for A than the life of agent B for B, insofar as the object of A's desires, P_A , obtains whereas the object of B's desires, P_B , fails to obtain. *Prima facie* this argument seems very compelling, in that we seemingly have a grasp of what it is for there to be a difference between a state of affairs obtaining, and an agent merely *believing* that the state of affairs obtains when in fact it fails to obtain. Thus, where the *experience machine* thought experiment failed to provide sufficient justification for premise 7, i.e. that *state of mind theories* are implausible, the *parallel lives* thought experiment seems to succeed.

Despite the *prima facie* success of this argument, I shall argue that ultimately *state of the world* theories of a good life are less plausible than their *state of mind* counterparts. In doing so I intend to undermine the strength of the abovementioned argument derived from the *parallel lives* thought experiment.

Chapter 3

Contra State of the World Theories of a Good Life

Thus far we have formulated the *state of the world* theorists' argument against *state of mind* theories, and have presented the most plausible form of that argument, as well as the argument which has the most probability of succeeding, namely, the argument relying on the *parallel lives* thought experiment. In this section I intend to present direct argument against *state of the world* theories of a good life, and in doing so intend to undermine the *parallel lives* argument against *state of mind* theories of a good life.

3.1 Subjectivity and Objectivity

Related to, but not identical with, the *state of mind/state of the world* distinction is a further distinction between theories of a good life, namely, the distinction between *subjective* theories of a good life, and *objective* theories of a good life.

According to Sumner, the subjective "is that which pertain to, or is characteristic of subjects" where a *subject* is "anything capable of conscious states of processes".²¹ Consciousness is "understood *very broadly*" so as to include all cognitive, conative, and affective states and processes (e.g. thought, deliberation, perception, sensation, emotion, memory, desire, imagination, dreaming, appetite, etc.).²² All agents are subjects insofar as all agents are capable of cognitive, conative, and affective states

²¹Ibid., 27.

²²Ibid., *italics added*.

and processes; though, presumably, not all subjects are agents, insofar as we would classify certain animals as subjects but reserve the category of agents for persons. What differentiates agents from subjects is beyond the scope of this thesis, thus we need only keep in mind that all agents are subjects. The subjective, in as much as it is that which pertains to a subject, where a subject is anything capable of cognitive, conative, and affective states and processes, is that which pertains to cognitive, conative, and affective states and processes. Subjectivity is thus mind-dependent, inasmuch as it pertains only to subjects, i.e. those things capable of mental (cognitive, conative, and affective) states and processes, i.e. those things which have minds.²³ Likewise, if some state or process is mind-dependent, it also subjective, since it pertains to subjects, i.e. those things capable of mental states and processes.²⁴

Having defined *subjective* in the above way, where the subjective is mind-dependent, we may in turn define the *objective* as *not* mind-dependent; such that some state or process is objective if it is not a mental state or process. On this view, then, a state of affairs P is objective if and only if it is not a mental state M , such that $P \neq M$. Conversely, a state of affairs P is subjective if and only if it is a mental state M , such that $P = M$.

If whether or not an agent A_i 's life is a good life is 'subjective' we would expect that the A_i 's mental (cognitive, conative, and affective) states M_i would necessarily figure into any evaluation of the life of that agent; that is, we would expect whether or not that agent's life is a good life for that agent to be mind-dependent, dependent

²³Ibid.

²⁴Ibid.

upon that agent's cognitive, conative, and affective states.²⁵ Conversely, if whether or not an agent A_i 's life is a good life is 'objective' we would expect that the mental states M_i of agent A_i would not factor into the evaluation of whether A_i 's life is good for A_i ; that is, we would expect that whether or not that agent's life is a good life for that agent to not be mind-dependent, not dependent on A_i 's mental states M_i . On an *objective* theory of a good life it could be the case that an agent A_i 's life is a good life for A_i regardless of A_i 's cognitive, conative, or affective states, that is, regardless of what A_i thinks, feels, wants, desires, wishes, etc. On such a view the life of an agent A_i could be a good life for A_i even if A_i is *miserable*. Though, surely, it is absurd to think that the life of an agent A_i could be good for A_i irrespective of the sort of attitudes that A_i takes toward their own life. As such, an *objective* theory of a good life is absurd, therefore, the only viable theories of a good life are *subjective* theories.

The rejection of *objective* theories of a good life rule out certain formulations of *state of world* theories of a good life; namely, W_2 , W_4 , W_6 , and W_8 , those are the formulation of *state of the world* theories which include an 'if and only if' clause. The remainder, W_1 , W_3 , W_5 , and W_7 , are all open to *subjective* interpretation, insofar as they do not deny that states of mind *could* have an impact on an agent's life, they merely assert that states of the world *do* have an impact. Thus, any *state of the world* theory is viable in as much as it also permits for the possibility that mental states somehow figure into the evaluation of an agent's life.

Sophisticated proponents of *state of the world* theories recognize the difficulties involved in denying that mental states play any part in the evaluation of an agent's

²⁵Ibid., 37.

life. Beyond the example already given, of how a purely objective theory might result in the claim that an agent A_i has a good life even if A_i is miserable, additional difficulties present themselves to those who would deny that mental states play at least some role in determining whether or not an agent has a good life.

3.1.1 Problems for Objective State of the World Theories

The first difficulty that emerges for *objective state of the world* theories of a good life is that were some *objective state of the world* theory W^* of a good life true, then it would follow that if some state of affairs P_i (where P_i is *not* a mental state M_i , and where P_i is the propositional content of agent A_i 's psychological states) obtains *after* A_i is already dead, then A_i 's life would be made better due to P_i obtaining. To give a more concrete example, if the work of Bertrand Russell on nuclear disarmament were to today result in developments which yielded further nuclear disarmament, then Bertrand Russell's life would be a better life for Bertrand Russell than if his life's work for nuclear disarmament had no such impact.²⁶ It would follow from W^* that the lives of deceased agents would be made better off if states of affairs (which were the propositional content of these agents' psychological states) were to obtain after the agents were already deceased. In short, the lives of the dead would be impacted by things which could not possibly affect their mental states, insofar as there would be no mental states to affect. This is a very strange and implausible conclusion.

Griffin attempts to salvage this position by stating that "there seems nothing

²⁶Griffin, 22-23.

irrational in attaching value to [...] posthumous [aims]”.²⁷ This sort of defense of the claim that the dead can have good lives misses the point. There is indeed nothing irrational about a *living* agent attaching value to posthumous aims, such as an agent who desires that their children’s lives be good lives even once they themselves are deceased, nevertheless, the claim of W^* is much stronger. It asserts that the life of an agent A_i can be impacted by states of affairs P_i which obtain after A_i ceases to exist. Since when we evaluate whether or not the life of an agent A_i is a good life, we evaluate whether it is a good life for A_i , if there is no agent A_i then there can no longer be a life which is good for A_i . Which is not to say that we cannot posthumously evaluate the life of an agent, rather only that we cannot posthumously evaluate the life of an agent for the agent whose life it was, since there is no longer an agent to speak of.

The second difficulty which exists for any *objective state of world* theory W^* of the good life involves considering a states of affair P_i (where P_i is the propositional content of some psychological states of agent A_i) which obtain even though A_i fails to be aware of P_i obtaining. If it is the case that W^* is true then it would follow that some state of affairs P_i , which is not a mental state M_i , could obtain without A_i being aware that P_i , such that P_i obtaining makes A_i ’s life better for A_i . That is, the fact that P_i obtains, even though A_i is unaware of this fact due to either special or temporal distance, would nevertheless make A_i ’s life good for A_i .

Consider the following example. Suppose that there is some agent A who desires that humans make contact with extraterrestrial life, that is, the state of affairs which is the propositional content of A ’s desire is that ‘humans make contact with

²⁷Ibid., 23.

extraterrestrial life':²⁸ let this state of affairs which is the propositional content of A's desire be P. Unknown to A humans have already made contact with extraterrestrial life, it is already the case that P, it just so happens that, as per some conspiracy theories, our governments are keeping this information from the general public, as such A is unaware of that P, and believe that $\neg P$. If W^* is true, then it would have to be said that because A's desire is, in some sense, satisfied, insofar as the object of A's desire P obtains, and, as such, that A has been made better off. Nevertheless, it seems strange to say this when we consider A's psychological state in contrast with some alternate agent A^* , who is psychologically indistinguishable in all respects from A, except who does know that we have made contact with extraterrestrial life, i.e. where A^* is aware of the fact that P. Presumably A^* 's psychological states, if we could examine them, show a greater level of satisfaction, relative to A. Moreover, let us consider a third agent A^{**} , who *falsely believes* that we have made contact with extraterrestrial life, when in fact we have not, such that it is the case that P, but A^{**} believes that $\neg P$. A^{**} 's psychological states would, presumably, be identical with those of A^* , despite A^{**} 's states being brought about by *false beliefs*. If W^* were true it would be the case that A's life is as well off as that of the A^* , despite the fact that A remains unaware of the object of A's desire obtaining, despite the fact that A remains unaware that it is the case that P. And moreover, it would also be the case, if W^* were true, that the life of A^{**} would be worse off than that of the first two agents, despite it being psychologically indistinguishable from that of agent A^* . In short, we would be forced to conclude that the life of the A, who fails to realize that the object of his desires obtains is better off than the life of the A^{**}

²⁸Sumner, 125.

who *falsely believes* that the object of his desire obtains despite the fact that A *feels* less satisfied than A**.

3.2 The Experience Requirement

Given the abovementioned problems *objective state of the world* theories of a good life face, that is, *objective state of the world* theories result in highly counterintuitive conclusions about when an agent's life is a good life. As such, any viable *state of the world* theory of a good life will be a *subjective state of the world* theory of a good life, a theory which takes into consideration mental states, as well as states of the world, in its evaluation of the lives of agents. This fact has been realized by proponents of *state of the world* theories of a good life such as Griffin and Sumner. In having realized the absurdity of the *objective* theories, Griffin and Sumner seek to preserve *state of the world* theories by incorporating an *experience requirement* into *state of the world* theories. The *experience requirement* states that for a state of affairs P_i to affect the life of an agent A_i , P_i must experientially interact with A_i : that is, A_i must experienced P_i .²⁹ Let us render a state of affairs P_i which has been experienced by agent A_i as $E_i(P_i)$.

The inclusion of the experience requirement, while making the *state of the world* theories *subjective*, allegedly does not collapse *state of the world* theories into *state of mind* theories insofar as *state of mind* theories “treat veridical and illusory experiences which are phenomenologically indistinguishable as equally valuable” whereas on any *state of the world* theory “the actual occurrence of the desired state of affairs

²⁹Griffin, 16.

is one necessary condition” in the evaluation of whether or not an agent’s life is a good life for that agent.³⁰

The inclusion of the *experience requirement* into *state of the world* theories, that is, making *state of the world* theories *subjective*, supposedly allows them access to the best of both worlds. These theories are alleged to have all the desirable features of *state of the world* theories without their drawbacks, as well as all the desirable features of *state of mind* theories without their drawbacks. A *subjective state of the world* theory W_{\bullet} avoids the absurd conclusions of W^* , in that on W_{\bullet} it does not follow that a state of affairs P_i of which an agent A_i is unaware of, that is $\neg E_i(P_i)$, can have an impact on A_i ’s life, nor that states of affairs can have an impact on A_i ’s life once A_i is deceased, since, once again, any state of affairs P_i which occurs after A_i is deceased cannot experientially interact with A_i , as such any such state of affairs P_i will be $\neg E_i(P_i)$. Likewise, since W_{\bullet} is a *state of the world* theory, it would also follow that if some state of affairs P_A , which is the propositional content of some cognitive, conative, or affective state of A , does not obtain where A is unaware of $\neg P_A$, that is where $\neg E_A(\neg P_A)$, that agent A does not have an equally good life to some agent B , who is psychologically identical to agent A , except that P_B actually obtains where $E_A(P_B)$. Thus W_{\bullet} is supposedly sensitive to whether or not an experience is veridical or illusory, unlike *state of mind* theories.

3.2.1 The Failure of the Experience Requirement

Despite the alleged advantages of incorporating an *experience requirement* into *state of the world* theories, it turns out to be the case that the incorporation of the

³⁰Sumner, 128.

experience requirement results in a self-defeating theory: any such theory is self-defeating in the sense that the *state of the world* theory which incorporates the *experience requirement*, contrary to what Sumner claims, collapses into a *state of mind* theory, given that *state of the world* and *state of mind* theories are mutually exclusive.

On the *experience requirement* a state of affairs P_i has an impact upon the life of agent A_i if and only if P_i experientially interacts with A_i , i.e. if and only if A_i experiences P_i , i.e. iff $E_i(P_i)$. In other words, if $\neg E_i(P_i)$ then the state of affairs P_i does not have any impact upon the life of agent A_i . The negation of the state of affairs P_i , that is $\neg P_i$, is itself a state of affairs. As such, given the *experience requirement*, if $\neg E_i(\neg P_i)$ then the state of affairs $\neg P_i$ does not have any impact upon the life of agent A_i . W_\bullet is suppose to be able to be sensitive to whether an experience is illusory or veridical, that is, as a *state of the world* theory, in the *parallel lives* thought experiment, it is suppose to count the life of agent A, who desire P_A , where $\neg P_A$, where A believe that P_A , as less good for A than the life of agent B, who desire P_B , where P_B and B is aware of P_B , is good for B; that is, as a *state of world* theory it is suppose to yield the result that life of A is less good for A than the life of B is for B, such that the life of A < the life of B.

Nevertheless, given the *experience requirement*, W_\bullet cannot yield this conclusion. Consider the following:

- 1 Let some *state of the world* theory W_\bullet of the good life be true. [Assumption for *reductio ad absurdum*]
- 2 Then, given the statement of the state of world theory W_\bullet , it is *not the case*

that for any agent A_i there is a state of the world P_i , such that for agent A_i , A_i is having a good life *if* P_i obtains AND $\neg E_i(P_i)$. [from 1, given the characterization of W_\bullet]

3 Also, given the statement of the state of world theory W_\bullet , it is the case that for any agent A_i , there is a state of the world P_i , such that for agent A_i , A_i is having a good life *if* P_i obtains. [from 1, given the characterization of W_\bullet , specifically W_1]

4 Let persons A and B live psychologically (i.e. cognitively, affectively and conatively) identical lives. [Assumption]

5 Let P_A and P_B be states of the world such that:

(a) P_A is the propositional content of some cognitive, affective, or conative state of person A and P_B is the propositional content of some cognitive, affective, or conative state of person B, AND

(b) the cognitive, affective, or conative states of which P_A and P_B are propositional contents be in some way relevant to the question of whether A and B are having good lives. [Assumption]

6 Let it be the case that $\neg P_A$ without A knowing this, that is, where $\neg E_A(\neg P_A)$, but let it be the case that P_B where $E_B(P_B)$.

7 Then, given 3, the life of B is better than the life of A.

8 Also, given 2, A and B have equally good lives, since if it is the case that A is not aware that $\neg P_A$, then it follows that $\neg E_A(\neg P_A)$, and as such, by the

experience requirement (premise 2), $\neg P_A$ cannot have an impact upon the life of A. Hence, contrary to what a *state of the world* theory would have to say, if the experience requirement is true, then the fact that P_A does not obtain, where P_A is the propositional content of some desire of A, that is, the fact that $\neg P_A$ obtains, where it is the case that $\neg E_A(\neg P_A)$, $\neg P_A$ obtaining does not and cannot impact the life of A.

9 Therefore, given 7 and 8, it is the case that both the life of A = the life of B, in terms of how good the lives are for their respective agents, AND that the life of A < the life of B.

10 BUT, 9 is absurd.

11 Therefore, by *reductio*, 1 is false.

Where we may substitute premise 2 with one of the following:

2A Then, given the statement of W_\bullet , it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is *not* having a good life *if* P_i obtains AND $\neg E_i(P_i)$.

OR

2B Then, given the statement of W_\bullet , it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if and only if* P_i obtains AND $\neg E_i(P_i)$.

OR

2C Then, given the statement of W_{\bullet} , it is *not the case* that there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if* P_i obtains AND $\neg E_i(P_i)$.

OR

ETC.

Likewise, we may substitute 3 with one of the following:

3A Then, given the statement of W_{\bullet} , it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is *not* having a good life *if* P_i obtains.

OR

3B Then, given the statement of W_{\bullet} , it is *not the case* that for any agent A_i there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if and only if* P_i obtains.

OR

3C Then, given the statement of W_{\bullet} , it is *not the case* that there is a state of the world P_i , such that for that agent A_i , A_i is having a good life *if* P_i obtains.

OR

ETC.

With each of the respective substitutions for premise 2 and premise 3 we cover all possible formulations of *subjective state of the world* theories. Since the premises which we are substituting are relevantly similar to premise 2 and premise 3, the conclusion likewise follows from any of the substituted premises.

3.3 Implausibility of State of the World Theories

Therefore, the most plausible form of *state of the world* theories, the form which incorporates the *experience requirement*, turns out to be inconsistent, insofar as the incorporation of the experience requirement into a *state of the world* theory collapses it into a pure *state of mind* theory. Hence, while it is the case that *objective state of the world* theories are implausible due to the counterintuitive conclusions which they force us to draw, *subjective state of the world* theories collapse into *state of mind* theories. Therefore, the choice that we are presented with seems to be either *objective state of the world* theories or *state of mind* theories.

The conclusions of *objective state of the world* theories, namely, that dead agents can have good lives, as well as that states of affairs which do not enter into the experiences of an agent can impact the life of that agent, are far more implausible than the conclusions that *state of mind* theories draw. As such, as it stands, *state of mind* theories, as counterintuitive as its conclusions might be, are nevertheless less implausible than their only viable *state of the world* counterparts.

We might formulate a direct argument against *state of the world* theories as follows:

- 1 Either it is the case that some *objective state of the world* theory W^* of the good

life is true OR it is the case that some *state of mind* theory M^* of the good life is true.

- 2 Let persons A and B live psychologically (i.e. cognitively, affectively and conatively) identical lives. [Assumption]
- 3 Let P_A and P_B be states of the world such that:
 - (a) P_A is the propositional content of some cognitive, affective, or conative state of person A and P_B is the propositional content of some cognitive, affective, or conative state of person B, AND
 - (b) the cognitive, affective, or conative states of which P_A and P_B are propositional contents be in some way relevant to the question of whether A and B are having good lives. [Assumption]
- 4 Let it be the case that $\neg P_A$ without A knowing this, but let it be the case that P_B .
- 5 Then, if M^* is true (given the statement of the state of mind theory M^*), despite 4, both A and B are having equally good lives.
- 6 BUT, 5 is not plausible.
- 7 Also, if W^* is true (given the statement of the objective state of mind theory W^*) it is the case that for any agent A_i there is a state of the world P_i , such that for agent A_i , A_i is having a good life if P_i obtains AND $\neg E_i(P_i)$.
- 8 If 7, then it is the case that a state of the world P_i can affect the life of a dead agent A_i .

9 Also, if 7, then it is the case that a state of the world P_i can affect the life of a agent A_i even if A_i does not experience P_i , i.e. even if $\neg E_i(P_i)$.

10 BUT, 8 and 9 are not plausible.

11 Furthermore, 8 and 9 are *more* implausible than 5.

12 Therefore, M^* is more plausible than W^* .

Noting, of course, that we can substitute premise 7 with some suitable reformulation thereof while still maintaining the validity of the argument.

Hence, even granting that some *state of mind* theory M^* of the good life is implausible, it is still the case that M^* is more plausible than some *state of the world* theory W^* . Such that, for all the failings of M^* , it is a better theory of a good life than W^* .

Chapter 4

Defense of State of Mind Theories

Having established that any *subjective state of the world* theory W_\bullet , that is, any *state of the world* theory which incorporates the *experience requirement*, collapses into a *state of mind* theory, also having argued that any *objective state of the world* theory W^* is more implausible, and minimally as implausible, as any *state of mind* theory M^* , the situation now is such that both W^* and M^* are on roughly equal footing. Both theories fail to adequately account for our intuitions concerning good lives. As such, in order to further claim that some *state of mind* theory M^* of a good life is superior to any *state of the world* theory W^* of a good life further arguments must be provided as to why we should prefer M^* over W^* .

4.1 States of Mind Always Matter

Setting aside the problems involved in attempting to integrate an *experience requirement* into *state of the world* theories, ultimately, any plausible *state of the world* theory must be *subjective*, in the sense that it must *somehow* be connected to the cognitive, conative, and affective states of agents to be plausible.

Most proponents of *state of the world* theories would not deny this, after all their position is that states of the world *do* matter to the life of an agent, not that they are the *only* things which matter. As such, for any state of the world P_i , where P_i is the propositional content of some cognitive, conative, or affective state of agent

A_i , if P_i obtains and $E_i(P_i)$, that is, and A_i is aware that P_i obtains, A_i cannot be indifferent to P_i obtaining. Consider a more concrete example, if some agent A desires to visit Singapore, where visiting Singapore is the propositional content of A 's desire, if A does visit Singapore it would have to be the case that A regarded this state of affairs obtaining, the visit to Singapore, in a positive manner. If A was entirely indifferent to the state of affairs obtaining then presumably it wouldn't have any impact on whether or not A has a good life. Now, it might be the case that if a state of affairs P_i is the propositional content of some cognitive, conative, or affective state of agent A_i then this necessarily means that A_i is not indifferent to P_i , that is, A_i has some sort of pro-attitude (or con-attitude) toward P_i . If it were the case that A_i lacked this appropriate pro-attitude (or con-attitude) toward P_i then presumably P_i would not be the propositional content of some cognitive, conative, or affective state of A_i , and, as such, A_i would be indifferent to whether or not P_i obtains. For example, if I am indifferent to the outcome of the next US presidential elections, and, moreover, the outcome of the US presidential elections does not have any impact upon any of the states of affairs which are the propositional contents of my cognitive, conative, and affective states, then no matter what the outcome of the US presidential elections my life goes equally well for me. As such, any plausible *state of the world* theory ultimately also depends upon states of mind in its evaluation of the life of an agent for that agent. If an agent is indifferent to some state of affairs obtaining, and, furthermore, that state of affairs obtaining has no impact upon any states of affairs which the agent does care about, then the state of affairs which the agent is indifferent toward has no impact upon the life of the agent.

This only establishes that states of mind *do* matter, not that they are the only

things that matter, the latter being the claim which must be shown if *state of mind* theories of the good life are correct.

4.2 Why do we think States of the World matter?

The argument derived from the *parallel lives* thought experiment exposes the fact that most people do believe that states of the world matter to whether or not our lives are good for us. I shall here attempt to provide an explanation as to why we believe this, even if it is false, and in doing so intend to undermine the intuition that underlies the *parallel lives* argument against *state of mind* theories of a good life.

4.2.1 States of the World Matter Most of the Time...

The *prima facie* appeal of the *parallel lives* argument, the argument that states of the world must matter to our lives, is derived from the simple fact that in *most* circumstances states of the world *actually do matter*. Nevertheless in carrying over this intuition to cases such as those of the experience machine we step outside of the bounds of *most* circumstances, such that the circumstances that an agent is in inside an experience machine are sufficiently different and extraordinary as to make our intuitions mistaken.

Under most circumstances, that is, in ordinary life, an agent A_i possesses verification conditions for whether or not some state of affairs P_i obtains or fails to obtain; A_i can test for whether or not P_i actually obtains or not. That is to say, if A_i happens to have a false belief regarding some state of affairs P_i , A_i can, in principle, verify or falsify this belief. The state of affairs P_i , or $\neg P_i$, is capable of

entering into agent A_i 's field of experience in some manner as to affect A_i 's beliefs about P_i , and thereby it can impact whether or not A_i has a good life or not. For example, under normal conditions, if I am being deceived by the people around me such that they only pretend to like me when in reality they detest me, it is unlikely that their deception is so perfect and elaborate as to prevent me from finding out that they detest me; someone is likely to slip up and the truth will come out. I, of course, would be devastated by such a revelation. Hence, in this case, because the state of affairs of the people around me detesting me is capable of entering into my experiences it is capable of impacting my life, thereby making it worse for me. Thus, naturally, the state of affairs in question in that example *does* matter to whether or not I have a good life. Still further, not only is it the case that states of affairs in most circumstances *can*, in principle, enter into our experiences, such that it is, in principle, possible to either verify or falsify our beliefs about said states of affairs, it is also the case that in *most* circumstances states of affairs *do* enter into our field of experiences whether we like it or not. For example, if it is the case that the people around me detest me it is unlikely that they would go about putting up an elaborate facade to make me believe that they actually like me, it is far more likely that they would show this detest outright, such that I would know right away whether they detest me or not.

Consider a thought experiment involving an agent who desires to truly be liked by the people around them, and believes that they are indeed so liked, where this belief happens to be *false*; those around the agent in question actually happens to detest them, they merely put on an elaborate facade in order to convince the agent that they like them, but behind the agent's back they slander them and reveal their

true feelings. In the case of thought experiments like this one, thought experiments designed by philosophers, we may stipulate that the agent in question *does not* find out that their belief is *false*, nevertheless, since in ordinary circumstances it is not the case that we can guarantee that an agent will never find out that their belief is false, this might lead us to unintentionally carry over the suspicion to the thought experiment itself. Sooner or later in many ordinary real life cases the person who is the victim of the deception comes to find out that their friendships are false, and that the people they care about, and wish would care about them, do not care about them. Likewise, for most of the objects of our desires, it is relatively easy to check for whether or not the objects of our desires actually obtain or not, most of our desires are not so fanciful as to be incapable of verification or falsification. Thus, it is possible that when we evaluate the life of an agent with *false beliefs* about the objects of their desires, and evaluate them not on the basis of that agent's occurrent experience of their life, what we unintentionally do is discount the value of that life for that agent on the basis of the *possibility*, or even *likelihood*, that the agent will find out that their belief is false. Since, under normal circumstances and with regards to most desires, it is very much possible that the agent will, sooner or later, themselves come to discover that their beliefs are *false*, we, as observers with information which the agent in question presently lacks, evaluate that agent's life *as though* the agent had the relevant information.

Moreover, given that when an agent A_i desires some state of the world P_i , they presumably desire that the state of the world P_i obtain, not that they merely have the experience of P_i obtaining, that is, not merely that $E_i(P_i)$, under normal circumstances A_i would seek confirmation of P_i having obtained by attempting to somehow

experience P_i . The only means by which we have access to the states of the world is via our experiences. As such, for example, if some person were informed that a close childhood friend with whom they have not had frequent contact with, but who they still care about, is getting married by a source that they consider unreliable, rather than immediately being happy for the childhood friend, they would seek to first confirm this claim by getting in contact with the childhood friend. Such an agent would be attempting to ground their beliefs in a state of affairs via experience. And, as such, states of affairs *can* and *do* interact with an agent experientially they matter to the life of the agent.

From these facts those who subscribe to *state of the world* theories draw the erroneous conclusion that because states of the world matter *most* of the time, they *must* matter *all* the time.

4.2.2 ...but NOT Always

The experience machine is presumably a perfect simulation, such that it is entirely indistinguishable from the real world; there is no experience that could be had in the experience machine which would give away that it is a simulation. Hence, inside the experience machine no verification conditions exist for finding out that all of the experiences in the machine are simulated; that is, an agent A who is inside the experience machine cannot experience anything which will make him believe that any of his experiences are illusory. Therefore, the simulated nature, that is, the illusory nature, of the experiences inside the experience machine could never enter into the field of experiences of an agent inside the machine. As such, some state of the world P_i can never have any impact upon the mental states of an agent A_i inside

the experience machine.

But having already argued that any plausible *state of the world* theory of a good life must take into consideration the mental states of agents, that is, that it must be subjective, it follows from this that for any agent A_i that is inside the experience machine, there will be no state of the world P_i which will affect the mental states of A_i . Whether or not P_i obtains will have no impact upon the mental states of A_i . Such that if any *state of the world* theory claims that P_i impacts the life of A_i it would be violating subjectivity, since subjectivity is concerned with the mental states of agents, and since P_i cannot have an impact upon the mental states of A_i , the claim that P_i affects the life of A_i would be an objective claim, and, in virtue of this, could not be made by any *subjective state of the world* theory of a good life.

We might formulate the arguments as follows:

- 1 The most plausible form of a *state of the world* theory is some *subjective state of the world* theory $W\bullet$.
- 2 Any *subjective* theory must take into considerations the cognitive, conative, and affective (i.e. mental) states of an agent.
- 3 Thus, from 1 and 2, if some *subjective state of the world* theory $W\bullet$ of the good life is true then any evaluation of the life of agent A_i for A_i must take into consideration the mental states of A_i .
- 4 Let persons A and B live psychologically (i.e. cognitively, affectively and conatively) identical lives. [Assumption]
- 5 Let P_A and P_B be states of the world such that:

(a) P_A is the propositional content of some cognitive, affective, or conative state of person A and P_B is the propositional content of some cognitive, affective, or conative state of person B, AND

(b) the cognitive, affective, or conative states of which P_A and P_B are propositional contents be in some way relevant to the question of whether A and B are having good lives. [Assumption]

6 Let it be the case that $\neg P_A$ without A knowing this, but let it be the case that P_B .

7 $\neg P_A$ cannot have an impact upon the mental states of agent A when $\neg E_A(\neg P_A)$.
[given 3]

8 $\neg E_A(\neg P_A)$. [given 5]

9 Thus, $\neg P_A$ cannot have an impact upon the mental states of A. [give 6 and 7]

10 Thus, that $\neg P_A$ has no impact upon the life of A. [given subjectivity]

11 Thus, both A and B are having equally good lives.

Therefore, even on the most plausible *state of the world* theory W_\bullet of a good life it turns out to be the case that A and B have equally good lives. As such, the most plausible form of *state of the world* theories yields the same conclusion as the most plausible form of *state of mind* theories.

Our initial reluctance to accept the conclusion that agent A and agent B live equally good lives, despite the experiences of A being illusory and those of B veridical, can thus be explained away by appeal to the fact that under *normal* circumstances

the experiences of an agent are verifiable or falsifiable against the states of affairs. But since the agent A inside the experience machine is in *abnormal* circumstances the intuition that we have with regard to states of affairs mattering to the life of agent A no longer apply, as the only means by which a state of affairs can impact an agent's life is experientially, something that *subjective state of the world* theories accept, it follows that since the state of the world cannot experientially impact the life of A, the state of the world does not matter in the evaluation of the life of agent A.

4.3 The Agent's Own Evaluation of Their Life

If it is the case that when we evaluate the life of an agent as a good life what we are doing is evaluating the life of that agent for the agent whose life it is, that is, the life of A for A, and presumably not for some other agent B, then A's own evaluation of their life should carry some weight in our evaluation of A's life. In the *parallel lives* thought experiment, since agent A and agent B are psychologically identical, they will presumably also give identical evaluations of their lives. Furthermore, we have no reason to doubt the genuineness, or authenticity, of the evaluation of either agent. That is to say, presumably both agent A and agent B are being completely honest both with us and with themselves when they evaluate their respective lives as good. How is it then, that we, as third party agents, can override the agent's own evaluation of their life?

An agent's evaluation of their own life is genuine if it reflects the *actual* attitude that agent takes toward their life. The fact that the experiences which an agent

is undergoing are illusory is no reason to doubt the genuineness of that agent's evaluation of their experiences.

To argue that were the agent aware that the experiences were illusory is no longer to evaluate that agent's life at time t , where t is now, and in circumstances C , where C are the circumstances in which the agent is in. Instead it is to evaluate the life of the agent in some alternate circumstances C^* , where the agent in question is presumably aware of the illusory nature of the experiences which they are undergoing. But this sort of evaluation presumes, contrary to what is suppose to be assumed in the *experience machine* thought experiment, that A is aware of the illusory nature of the experiences in the experience machine, which, *ipso facto* of the experience machine, A cannot be aware of. Hence, this sort of evaluation does not evaluate the life of A as it is, in circumstances C , but as it would were A aware of additional facts of which A is not and cannot be aware of.

Thus we cannot appeal to some *hypothetical* evaluations of agent A as indications that A 's life *here and now* as it *actually* is fails to be not a good life for A , insofar as the experiment machine makes it impossible for the *hypothetical* circumstances, in which A is aware of the illusory nature of the experiences, to take place. Such an appeal to agent's *hypothetical* evaluations of their own life has no impact upon the *actual* evaluation of A 's life for A . That is to say, we will well grant to the *state of the world* theorist that insofar as agent A prefers veridical experiences to illusory ones, were A aware of the fact that all their experiences were illusory naturally A would evaluate their own life as a bad life, A 's life would not be good for A ; nevertheless, since it is *not* the case and *cannot* be the case that A is aware of the illusory nature of the experiences in the experience machine, by the very nature of the experience

machine, the *hypothetical* evaluation fails to evaluate the *actual* life of A, but instead evaluates some *hypothetical* life.

Thus, the only way to reach the conclusion that A's life is not as good as B's life is to evaluate the life of A from the perspective where A is aware of the illusory nature of the experiences. But to do so is to violate the conditions of the experience machine thought experiment, and is as such invalid. It is not a proper evaluation of A's life inside the machine.

4.4 The Simulation and Skepticism

Another way to approach this experiment is not from the perspective of an agent who is asked to hook up to the experience machine after having lived in the real world, but instead from the perspective of agents who have always existed inside a simulated reality. Suppose that we have a community of agents who were actually very advanced artificial intelligences existing inside a simulation of the world. The simulation is identical in every experiential respect to our own world, which is to say, it is experientially indistinguishable from the real world, such that were anyone from the real world to be uploaded into the simulation they would be unable to tell the difference between the two worlds. Furthermore, our artificial agents would also be unaware that their world is a simulacrum of ours.

In this simulated world, let us call it S^* , in contrast with the real world, S , there would be artificial agents who go about their lives; the simulation would include things like jobs, families, leisure activities, etc. Given the fact that the S^* is indistinguishable from S the states of affairs P_i which are the propositional content of

the cognitive, conative, and affective states of some artificial agent A_i^* in S^* would be identical with the states of affairs P_i , which are the propositional contents of the cognitive, conative, and affective states of some actual agent A_i in S . Furthermore, suppose that the two worlds are incapable of interacting with one another, such that those in S can observe those in S^* , but cannot change anything inside S^* , but that those in S^* have absolutely no access to S .

The lives of any agent A^* in S^* , on the *state of the world* theory, would have to be counted as less good than the lives of some parallel agent A in S . But this seems implausible, for how can something which cannot even in principle interact with the simulated world S^* impact in anything in S^* ? For the inhabitants of S^* , S^* is the only world, the real world, there exist for them no verification criteria to demonstrate otherwise. From the perspective of those of us in S , since we can observe that S^* is a simulation running on a computer, we have a criteria by which to distinguish S from S^* , something that is lacking for those from within the simulation S^* .

Part of the problem is that for an agent A^* in S^* there are no meaningful verification criteria for the proposition “ S^* is a simulation”, which is to say, any agent A^* in S^* faces the same sort of skeptical concerns as any agent A in S with regards to skepticism. We cannot know that we are not brains in a vat, that we are not inside a simulated reality of the sort like S^* , but, moreover, not only is it the case that we cannot know this, but we have no verification or falsification criteria for a statement of that sort. If the lives of agents in S^* are undermined by the fact that S^* is a simulated reality, at least from the perspective of agents in S , then likewise the value of our lives should be undermined by skeptical concerns that our reality is simulated, and hence that all our experiences are illusory. Now, obviously, this

is an absurd conclusion. No one thinks that simply because we might all be being deceived by Descartes' evil genius, or brains in the vat, or something of that sort, that this has any impact upon whether or not our lives are good lives. As such, why should we think, as the *state of the world* theorists wants us to think, that the lives of agent in S^* are undermined by similar skeptical concerns?

The same applies to an agent in the experience machine, they are, for all intents and purposes, in the same situation as agents in S^* , such that there are no meaningful verification or falsification criteria for the proposition "all my experiences are illusory" in the experience machine. The same, of course, applies to real life. We cannot discern whether the *totality* of our experiences are veridical or illusory. We may ask whether some given experience E is veridical or illusory by checking it against our other experiences. If I suspect that those around me disdain me, rather than like me, I can attempt to find out by perhaps spying on them to see what they say about me when they do not know that I am around, or such. There are, in principle, means by which we can discern whether one experience is veridical or illusory, and that is by checking it against other experiences. Nevertheless, the same cannot be done for the totality of all experience, for there is nothing against which we can check it. To say that experience is to be judged against *reality*, whatever that might be, doesn't help us in the slightest, for we have no access to *reality* beyond our experiences. The *totality* of experience is not subject to verification or falsification in the same manner that a single experience, or a set of experiences, is, insofar as there is nothing against which we are capable of checking the veridical or illusory nature of the *totality* of experience.

Thus, not only is whether or not S^* is a simulation, or whether or not the expe-

periences of an agent A in the experience machine are illusory or veridical, something which cannot enter into the experiences of an agent A* in S*, or an agent in the experience machine, but it is not even intelligible to say that the *totality* of the experiences of agent A* or A could be veridical or illusory for A* or A.

When we evaluate the lives of agent A* in S, or agent A in the experience machine, we must be cognizant of the fact that what we seem to be doing is devaluing their lives based on purely skeptical concerns. If we were to attempt to put ourselves in the perspective of agent A* or A we would realize that they stand in the same relation to their experiences as we do to ours. We are in no better of a position to claim that our experiences are veridical than they are to claims that theirs are. As such, if such skeptical concerns are to undermine the value of the lives of A* and A for A* and A respectively, then likewise the same sort of skepticism undermines the value of our lives for us.

4.4.1 Truth and a Good Life

Another way of formulating the claims of *state of the world* theories is as follows: to say that what matters to the life of agent A_i is that some state of affairs P_i obtains, where P_i is the propositional content of some cognitive, conative, or affective state of A_i , is to say that it matters to the life of A_i that the proposition of which P_i is the content of be true. Let O_i be the proposition of which P_i is the content of, if some *state of the world* theory W^* of a good life is true then it is the case that the truth-value of O_i will affect the life of A_i , such that if O_i is TRUE then the life of A_i will be a good life for A_i , and, conversely, if O_i is FALSE, then the life of A_i will not be a good life.

Furthermore, a proponent of a *state of the world* theory will also claim that if some agent A_i is inside the experience machine it is the case that O_i is FALSE for A_i , where the propositional content of O_i is some state of the world P_i . *Prima facie* this seems plausible, nevertheless upon closer examination the claim is more dubious than it first seems.

Before we can say that O_i is false for A_i we must appeal to a particular theory of truth. Since it is not the case that on every theory of truth O_i will be false for A_i in the experience machine, the claim of the *state of the world* theorists depends upon a particular conception of truth.

If we take truth as *correspondence* then naturally it will be the case that O_i is false, insofar as the propositional content of O_i , i.e. P_i , fails to obtain. But, the correspondence theory of truth is itself suspect, insofar as there are alternative accounts of truth which would render O_i as true.

If, on the other hand, we take a *verificationist* or *falsificationist* view of the meaning of propositions, then the proposition O_i will be unverifiable or unfalsifiable for A_i , insofar as A_i is inside the experience machine, and there is no means by which A_i can either verify or falsify O_i . Likewise, the same goes if we take any other sort of *anti-realist* view with regards to truth, it could be the case on some *anti-realist* conception of truth that O_i is true for A_i . That is, that what it means for a state of the world P_i to obtain is merely that A_i experience P_i . If we take this *anti-realist* conception of *obtain* then it would follow that the two agents, A and B, in the *parallel lives* thought experiment would indeed have equally good lives, even on a *state of the world* theory, as it would be the case that for both A and B that P_i obtains.

Thus, setting aside all that has been said against *state of the world* theories thus

far, it still remains the case that whether or not A and B have equally good lives or not is dependent upon how we construe *truth* and how we construe a state of affairs *obtaining*. Since there are *anti-realist* conceptions of both *truth* and *obtaining* which even on a *state of the world* theory yield the conclusion that A and B have equally good lives, it is not a foregone conclusion that *state of the world* theories yield different results from *state of mind* theories. It must further be specified on which conception of *truth* and *obtaining* these theories yield the results which they do. It is, of course, beyond the scope of this thesis to defend any *anti-realist* conceptions of either *truth* or *obtaining*, but nevertheless we must be cognizant in these debates that there are indeed these different and competing conceptions which yield different results. Such that, the *state of the world* theorist who seeks to maintain, despite what has been argued in this thesis, that agent A and agent B, in the *parallel lives* thought experiment, do not have equally good lives, they have to further defend their *realist* conception of *truth* and *obtaining* since their conclusion depends upon these *realist* conceptions.

An advantage of a *state of mind* theory is that it does not have to concern itself with the *realism* vs. *anti-realism* debate. It is, in a way, a more elegant theory of a good life insofar as it does not depend upon contestable assumptions such as a *realist* or *anti-realist* conception of truth or of obtaining.

Chapter 5

Conclusion

Neither *state of the world* theories nor *state of mind* theories are without their share of difficulty. Both are plagued with certain problems, both result in certain unintuitive conclusions, in short, both are inadequate to fully capture all the intuitions which we have about good lives, as such, neither can properly account for the way in which we talk about good lives, that is, neither can fully account for the language of welfare. Nevertheless, despite the fact that both theories have problems it is not the case that both are equally plagued by problems. The major defect of *state of mind* theories, the seemingly counterintuitive conclusion that agent A and B, in the *parallel lives* thought experiment, have equally good lives, is a far smaller flaw than the problems faced by *state of the world* theories.

Firstly, *objective state of the world* theories, those *state of the world* theories which entirely ignore mental states in their evaluation of the life of an agent, produce far more problematic and counterintuitive conclusions than those of *state of mind* theories. If some *objective state of world* theory M^* of the good life is true then it would follow that states of affairs entirely outside of an agent's experiences can impact the life of an agent, and, likewise, that the life of an agent can be impacted by states of affairs which take place after the agent in question is already dead. On an *objective state of the world* theory M^* it is further the case that how the life of an agent *feels from the inside* for that agent is entirely irrelevant to the evaluation of that agent's life, hence, if some states of the world obtain which are regarded as

making an agent's life a good life, then that agent would have a good life *even if* the agent were *miserable*. All of these consequences of *objective state of the world* theories are patently unacceptable and absurd.

Secondly, *subjective state of the world* theories, those theories which attempt to incorporate mental states, alongside states of the world, in their evaluation of an agent's life, fail to get off the ground. In the flight from the absurdity of *objective* theories the *state of the world* theorists attempted to incorporate an *experience requirement* into *state of the world* theories so as to make them sensitive to the mental states of an agent. The *experience requirement* turned out to be overkill for their purposes, inasmuch as it resulted either in the resultant theory being patently inconsistent, or else collapsing into a *state of mind* theory. The incorporation of the *experience requirement* into a *state of the world* theory had the consequence of it yielding the result that agent A and B, from the *parallel lives* thought experiment, both had equally good lives, and that B's life was better than A's, thereby being inconsistent. Or, alternatively, it simply had the consequence of yielding that the lives of A and B were equally good for A and B respectively, and thus, while being sensitive to the mental states of an agent, it failed to be sensitive to the *state of the world*; it disregarded the *state of the world* in its evaluation entirely thereby collapsing into a *state of mind* theory.

Thirdly, and lastly, the *state of mind* theories of a good life could explain away why we might believe that its conclusion in the *parallel lives* thought experiment, that A and B have equally good lives, is counterintuitive. A *state of mind* theory, contrary to what its opponents claim, is indeed sensitive to the role that states of the world play in determining whether or not we live good lives. Insofar as

states of the world, in ordinary circumstances, experientially interact with agents these states of the world impact the mental states of an agent, and thereby impact the life of an agent. The effects on the lives of agents which states of the world have are mediated through mental states. The *experience machine* and *parallel lives* thought experiments represent a divergence from ordinary circumstances, they present us with very odd cases where an agent is entirely experientially isolated from the state of the world, seeing as we do not normally encounter such cases in *ordinary* circumstances we mistakenly carry over our intuition that states of the world impact our lives even in such *extraordinary* circumstances. Hence, a *state of mind* theory can account both for why it is the case that we have the intuitions that we do with respect to the relevant thought experiments, as well, it can also account for how it is the case that states of the world impact our lives, something which a *state of the world* theory seems incapable of doing.

Hence, contrary to what Nozick *et al.* give the impression of, *state of mind* theories are far less counterintuitive than their *state of the world* counterparts. The arguments presented here are not intended to be the final word on the subject; on the contrary, they are presented as challenge arguments. In order for a *state of the world* theory to be viable it must overcome the difficulties and arguments presented here. Furthermore, no specific *state of mind* theory was defended in this thesis, but rather *state of mind* theories in general, as such, there still remains the question of which particular *state of mind* theory is the correct theory, and connected to that, which states of mind are relevant in the evaluation of an agent's life; the question of which of *sensory hedonism*, *attitudinal hedonism*, some *preferentism* theory which regards desire fulfillment in a purely mental way, etc. is correct remains open.

Bibliography

- [1] Aristotle. *Nicomachean Ethics*.
- [2] David Brink. *Moral Realism and the Foundations of Ethics*. Cambridge University Press, Cambridge, 1989.
- [3] Stephen Darwall. “self-interest and self-concern”. *Social Philosophy and Policy*, 14:158–178, 1997.
- [4] James Griffin. *Well-Being: Its Meaning, Measurement and Moral Importance*. Clarendon Press, Oxford, 1986.
- [5] David Hume. *A Treatise of Human Nature*.
- [6] Jason Kawall. “the experience machine and mental state theories of well-being”. *The Journal of Value Inquiry*, 33:382–383, 1999.
- [7] Sumner. L.W. *Welfare, Happiness, and Ethics*. Clarendon Press, Oxford, 1996.
- [8] Robert Nozick. *Anarchy, State, and Utopia*. Basil Blackwell, Oxford, 1974.
- [9] Martha Nussbaum and Amartya Sen. *The Quality of Life*. Clarendon Press, Oxford, 1993.
- [10] Bernard Williams. *Moral Luck*. Cambridge University Press, Cambridge, 1980.