Life as a Trust Game

A Comment on *The Option Value of Life*

Gregory Ponthiere[1]

Abstract:
According to Burri (2020), a major reason why suicide is often irrational lies in the option value of life. Remaining alive is valuable because this allows for a larger menu of options, and the possibility of committing suicide in the future adds further value to the act of remaining alive now. In this note, I represent life as a trust game played by two selves – the young self and the old self – and I argue that the possibility to commit suicide in the future can encourage committing suicide now, against what the theory of the option value of life predicts.

Keywords: value of life, suicide, option value, trust game, multiple selves.

[1] UCLouvain, Hoover Chair in Economic and Social Ethics, 3 Place Montesquieu, 1348 Louvain-la-Neuve, Belgium.
E-mail : gregory.ponthiere@uclouvain.be URL : https://sites.google.com/site/gregoryponthieresite/home

1. Introduction

In a recent article, Burri (2020) uses the general theory of option value (Dixit and Pindyck 1994) to develop an original theory of suicide. In order to better understand the decision to remain alive or to commit suicide at a given point in time, Burri proposes to regard a suicide as an irreversible investment in non-existence, and to analyze the suicide decision from that perspective. According to that theory, a major reason why suicide is often irrational lies in the option value of remaining alive. Under the theory of option value, remaining alive is valuable, because this allows for a larger menu of options in the future, including the possibility to commit suicide if bad news turns out to occur.[2]

The theory of the option value of life tends to restrict the set of conditions under which suicide is rational: how can it be rational to commit suicide now, given that remaining alive still leaves open the possibility to commit suicide in the future in case of bad news? The option value of remaining alive measures the value of "waiting" before making an irreversible investment in non-existence. Under that theory, the possibility of committing suicide in case of bad news in the future contributes to add further value to the act of remaining alive.

In this comment, I argue, contrary to the theory of the option value of life, that the additional flexibility allowed by the possibility to commit suicide in the future does not necessarily make the act of committing suicide now less rational. The possibility to commit suicide in the future can encourage – rather than discourage – individuals to commit suicide right now, contrary to what the theory of option value predicts. Whether the extra degree of flexibility allowed by the possibility to commit suicide in the future makes current suicide more or less rational depends on the way in which one formalizes the suicide decision. Burri (2020) uses a particular framework: a single-agent rational choice model, where a unique agent weights gains and losses of his actions for the present and the future.

This paper develops an alternative behavioral model of suicide, based on *multiple selves* of the same person.[3] In that framework, the decision to commit suicide is modeled as a trust game played by two selves, the young self and the old self. The decision to remain alive or not is analyzed as a decision, made by a player (the young self), to invest or not in the existence of another player (the old self), within the trust game or investment game studied by Berg *et al.* (1995). Trust games are a particular class of sequential games, where a first player can decide either to "open the game" or to "close the game", while anticipating how the second player will play in the second stage. In this paper, the decision of the young self not to commit suicide is regarded as a decision to "open the game" (thus allowing for the existence of the old self), whereas the decision to commit suicide is interpreted as a decision to "close the game" (thus preventing the existence of the future self).

[2] The term "bad news" refers to the occurrence of an event that makes the decision-maker worse-off in comparison to the alternative case where "good news" occurs.

[3] Theories of multiple selves have become increasingly studied in philosophy and social sciences. In philosophy, the seminal reference is Parfit (1971). In economics, models of multiple selves have been developed to better understand saving decisions (Laibson 1997) and retirement choices (Diamond and Koszegi 2003).

Using this alternative framework, I show that the additional flexibility allowed by the possibility to commit suicide in the future can, under some conditions, encourage committing suicide now, in opposition to what *The Option Value of Life* predicts. Allowing for more flexibility in the future does not necessarily encourage remaining alive. The proof of that result is obtained by comparing two trust games that are exactly identical, except concerning the possibility of the old self to commit suicide in case of occurrence of bad news. It is shown that, under some conditions, the young self commits suicide when there is a possibility, for the old self, to commit suicide in the future, but would not commit suicide in the absence of such a possibility. Thus having more available options in the future does not necessarily discourage suicide now. Under some conditions, it may be quite the opposite.

The intuition behind that result goes as follows. Not committing suicide is a kind of investment made by the young self, and which allows the old self to exist. If the young self regards a future suicide of the old self as a lack of loyalty towards him (i.e. a lack of reward for his past investment in the existence of the old self), then the possibility, for the old self, of committing suicide can, in the absence of trust of the young self in the old self, encourage the young self not to invest in the old self, that is, to commit suicide. Hence more flexibility in the future can encourage the young self to "close the game" now, to avoid that the old self lacks loyalty towards him. That result can be interpreted by relying on the concept of commitment. The young self has an ideal plan for himself and for the old self. When there is a disagreement between the young self and the old self concerning the actions of the latter, the impossibility, for the old self, to commit to a particular action (e.g. no suicide) can encourage the young self, in the absence of trust in the old self, to "close the game", in order to avoid that the old self deviates from the ideal plan from the perspective of the young self.

2. Life as a trust game

Let us think about the life of a person as a trust game played by two selves of the same person: the young self and the old self (see Figure 1).[4] The young self plays first, and can decide to commit suicide (= "close the game") or not to commit suicide (= "open the game"). If the young self decides to commit suicide, the old self will not play, whereas if the young self decides not to commit suicide, the old self will play. If the old self can play, two situations can arise: the "good news" (with probability $1 - p$) and the "bad news" (with probability p). For simplicity, the old self will, in case of good news, necessarily not commit suicide, whereas, in case of bad news, he can either commit suicide or not. Figure 1 is a basic illustration of the theory of life as a trust game. Many real-life situations look like Figure 1, such as the case of an elderly person who will learn in the future (*via* a medical test) whether she is has a serious disease or not, and who decides to commit suicide now (or not). That decision can be modelled as a trust game played by her present self and her future self.

[4] Models of multiple selves of the same person are supported by experimental evidence. Regarding the saving decision, Hershfield *et al.* (2011) show that individuals interacting with realistic computer renderings of their future selves exhibit a stronger tendency to accept later monetary rewards over immediate ones.

To make a parallel with the seminal trust game introduced by Berg *et al.* (1995), this representation of life as a trust game amounts to regard, from the perspective of the young self, the strategy "no suicide" as an investment made by the young self in the old self, investment that allows for nothing less than the existence of the old self. On the contrary, the strategy "committing suicide" corresponds to the absence of investment of the young self, in the sense that, in that case, the old self will never come into existence. The suicide of the young self is regarded here as strictly equivalent to the non-existence of the old self.

The lives of the young self and the old self do not overlap. Obviously, the old self will be able to exist, and, hence, to make a decision, only if the young self does not commit suicide. But even if the old self does not exist materially at the time where the young self makes his decision, the decision made by the young self depends on his representations/beliefs concerning the behavior of the old self in case where he decides not to commit suicide, and, thus, to allow for the existence of the old self.

The pay-offs of the game are common knowledge to all players.[5] The pair (X, Y) has, as a first component, the pay-off to the young self, and, as a second component, the pay-off to the old self. Those pay-offs reflect how the two selves value the outcomes of the game.[6] Four outcomes are possible: (1) the young self commits suicide; (2) the young self does not commit suicide, good news occurs and the old self does not commit suicide; (3) the young self does not commit suicide, bad news occurs and the old self does not commit suicide; (4) the young self does not commit suicide, bad news occurs and the old self commits suicide.

The parameter $R_0 > 0$ is the pay-off to the young self in case of suicide. In the absence of suicide of the young self, the pay-off to the young self is equal to xR_0 when good news occurs in the future, to yR_0 when bad news occurs and the old self does not commit suicide, and to zR_0 when bad news occurs and the old self commits suicide. The pay-off to the old self is set to 0 when the young self commits suicide, and prevents the old self from existing.[7] If he exists, the old self enjoys a pay-off equal to aR_1 in case of good news, and to bR_1 or cR_1 in case of bad news, depending on whether the old self commits suicide or not.[8]

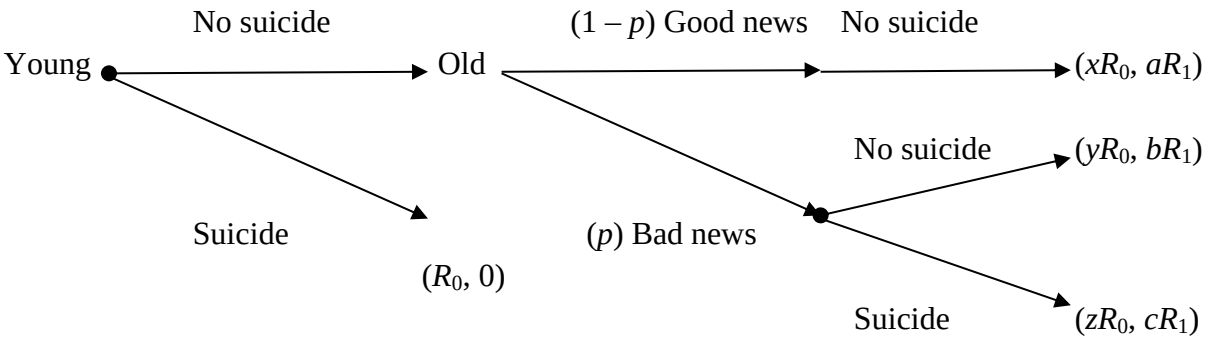
[5] It is assumed that the young self can anticipate the pay-off structure that will be faced by the old self, and that the old self can remember the plans that the young self had for the old self. Those assumptions amount to suppose some form of psychological connectedness between the temporal selves of the same person.

[6] Note that, since the young self and the old self are regarded as two distinct players of the game, there exists no preference ordering defined on the pairs of pay-off levels (X, Y). Such a preference ordering would amount to consider the existence of a "global self" taking the interests of the two selves into account and weighting these.

[7] 0 is a standard value for non-existence. Our results are robust to using another value Ω instead.

[8] R_1 is assumed to be strictly positive. No assumption is imposed on the relative levels of R_0 and R_1.

Figure 1: Game I (flexibility for the old self)



Parameters x, y and $z > 0$ capture how the young self values the outcome of the game in the case of no initial suicide. Given that the lives of the young self and the old self do not overlap, those pay-offs cannot be interpreted in hedonic terms. On the contrary, these represent how the young self evaluates, *at the time of his decision*, possible future outcomes related to (possible) decisions made by the old self. It is reasonable to suppose that the young self prefers the occurrence of good news to bad news, so that $x > y, z$, leading to a higher pay-off xR_0 under the good news than under the bad news (either yR_0 or zR_0).[9] Regarding parameters y and z, which capture the value that the young self assigns to the outcomes associated to the occurrence of bad news for the old self, three cases can arise. First, the young self is *indifferent* between the cases where the old self commits suicide or not in case of bad news: we have $y = z$. Second, the young self prefers the outcome where the old self commits suicide to the one where he does not commit suicide in case of bad news: we have $y < z$. Third, the young self prefers the outcome where, in case of bad news, the old self does not commit suicide to the one where he commits suicide: we have $y > z$.

Parameters a, b and c capture how the old self values the outcome of the game in the case of no suicide of the young self. It is reasonable to assume that the old self prefers the occurrence of good news to bad news, implying $a > b, c$. Moreover, we assume, for the sake of simplicity, that in case of bad news the old self prefers to commit suicide, that is, $c > b$.[10]

What is the outcome of this game? When deciding to commit suicide or not, the young self compares the pay-off $R_0 > 0$ in case of suicide with the pay-offs that arise if he does not

[9] We also impose $x > 1$. Otherwise, suicide would always be chosen by the young self, making the game trivial.
[10] That assumption is not crucial for our argument, which only requires the possibility of some disagreement between the young self and the old self regarding the desirable action of the old self in case of bad news. Thus, assuming $b > c$ would only lead us to illustrate our results by means of another game.

commit suicide, which are $xR_0 > 0$ in case of good news, and either yR_0 or zR_0 in case of bad news. The young self knows that the old self's best reply to the bad news is to commit suicide, since $cR_1 > bR_1$. Thus the expected pay-off, for the young self, from not committing suicide is: $(1 - p)\, xR_0 + p\, zR_0$. Hence the young self does not commit suicide if and only if:

$$(1 - p)\, xR_0 + p\, zR_0 > R_0 \quad \text{or} \quad (1 - p)x + pz > 1.$$

Given $x > z$, that condition is more likely to be satisfied when the probability of occurrence of good news $(1 - p)$ is large. Thus, as in *The Option Value of Life*, better prospects for the old self encourage the young self not to commit suicide.

Note that, as in any trust game, the first player may exhibit trust in the second player, and may believe that the latter will not choose what the best is for him, for the sake of respecting the plan of the first player. The young self could trust the old self, and could believe that the old self will not commit suicide in case of bad news, even though this would be the best option for himself. If so, the condition for no suicide of the young self becomes:

$$(1 - p)\, xR_0 + p\, yR_0 > R_0 \quad \text{or} \quad (1 - p)x + py > 1.$$

That condition is *weaker* than the above condition, since $y > z$. Trust of the young self in the old self makes the suicide of the young self less likely. On the contrary, a lack of trust of the young self in the old self encourages the former to "close the game", which coincides here with committing suicide, thus preventing the old self from existing. In that case, the young self would commit suicide to prevent the old self from deviating from his ideal plan.

3. Future options and the decision to commit suicide

According to the theory of the option value of life, the mere availability of the option of committing suicide in case of bad news in the future contributes to add further value to the act of remaining alive, and, hence, discourages suicide now. The goal of this section is to show that this is not necessarily the case in the model of multiple selves studied here. Actually, the possibility to commit suicide in the future does not necessarily make remaining alive more valuable. *Extra flexibility in the future can, under some conditions, have the opposite effect, that is, it can encourage suicide right now*.[11]

In order to provide a proof of that statement, this section compares the outcome of the game studied in Section 2 (where the old self can commit suicide in case of bad news) with the outcome of another game, which is exactly the same except that the old self cannot commit suicide in case of bad news. This amounts to delete, from the trust game studied

[11]Note that the comparison of the theory of the option value of life and the theory of life as a trust game played by multiple selves is relevant for the study of the effect of future flexibility on suicide decisions. Indeed, given the sequential nature of the trust game, there is, at any point in time, only *one* decision-maker who chooses to commit suicide or not, exactly as under the theory of the option value of life.

above, the branch of the tree corresponding to the suicide of the old self. That modified game (game II) is shown on Figure 2. Note that in game II, there is no disagreement among selves, and the issue of trust does not arise: the young self is sure that the old self will not commit suicide in case of bad news, and will thus act in line with what he regards as the best life-plan.

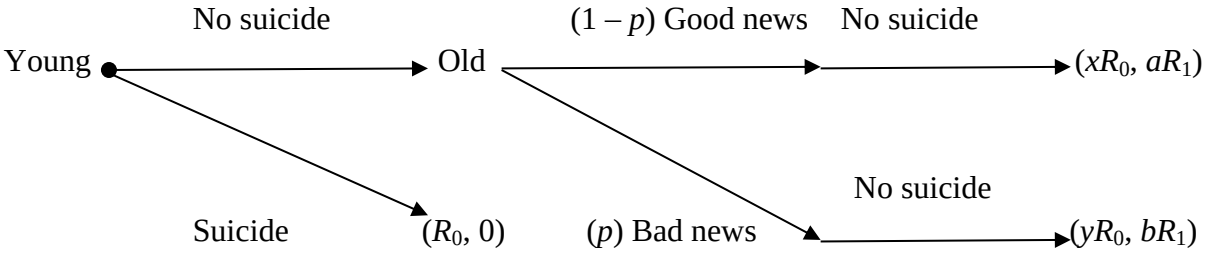
The comparison of trust games I and II can be used to show that it is not necessarily the case that additional flexibility (more options) in the future discourages suicide now. To prove this, we derive conditions under which the young self decides to commit suicide in game I (where the old self can commit suicide) but decides not to commit suicide in game II (where the old self cannot commit suicide).

In game II, the expected pay-off from not committing suicide is, for the young self: $(1 - p)\, xR_0 + p\, yR_0$. Hence the young self does not commit suicide if and only if:

$$(1 - p)\, xR_0 + p\, yR_0 > R_0 \quad \text{or} \quad (1 - p)x + py > 1.$$

If the young self prefers the outcome where the old self commits suicide to the outcome where the old self does not commit suicide (i.e. $y < z$) in case of bad news, the condition for no suicide from the young self is *stronger* in game II than in game I. Here the extra flexibility allowed by the possibility for the old self to commit suicide tends to encourage the young self not to commit suicide. This case coincides with the flexibility argument pointed out in *The Option Value of Life*.

Figure 2: Game II (no flexibility for the old self)



But this is only one case among three possible cases. Consider now the other cases.

If the young self is indifferent between the cases where the old self commits suicide or not (i.e. $y = z$), the no suicide condition in game II is identical to the one in game I (whatever

the young self exhibits trust in the old self or not). Hence, there is no effect of the possibility of committing suicide for the old self on the decision of the young self to commit suicide.

If the young self prefers the outcome where the old self does not commit suicide to the outcome where the old self commits suicide (i.e. $y > z$), the no suicide condition for the young self is *weaker* in game II than in game I in the absence of trust in game I, while the two conditions are identical when the young self trusts the old self in game I. To put it differently, if the young self dislikes the perspective of the old self committing suicide in case of bad news (i.e. he regards this future suicide as a lack of loyalty), and if he exhibits no trust in the old self, the set of probabilities of bad news for which the young self will commit suicide in game I is a superset of the corresponding set in game II. The extra-flexibility allowed by the possibility of suicide for the old self encourages the young self to commit suicide.

Let us show that the possibility to commit suicide in the future does not necessarily favor remaining alive now. To do so, it is sufficient to find a case where the young self commits suicide in game I but does not commit suicide in game II. Note that, if the young self exhibits trust in the old self in game I, the no suicide conditions in games I and II are exactly identical, so that this case cannot occur. Assuming the absence of trust in game I, the case where there is suicide of the young self in game I but not in game II occurs if and only if:

$$1 - py < (1 - p)x < 1 - pz$$

A necessary condition for those two conditions to hold is that the young self prefers that the old self does not to commit suicide in case of bad news ($y > z$), against the preferences of the old self (for whom suicide is more desirable). It is thus necessary, for having suicide in game I and no suicide in game II, that there exists a disagreement between the young and the old selves concerning the best action to be chosen by the old self.[12]

A disagreement between the young self and the old self in game I – together with the lack of trust of the young self in the old self – is thus necessary to have suicide in game I and no suicide in game II. But those conditions are not sufficient: we need also that there is a sufficiently large probability that bad news occurs, leading to the conflict of selves. Note that there exists a large set of values for parameters $\{x, y, z, p\}$ such that the two inequalities are satisfied, leading to suicide in game I and no suicide in game II.[13] Thus it is not true that additional flexibility (more options) in the future encourages the young not to commit suicide. This may actually have the opposite effect, as illustrated by the comparison of games I and II.

[12] The existence of such a disagreement does not necessarily occur. It could be the case, for instance, that the young self tends, by moral duty, to take the interests of the old self into account when valuing the outcomes of the game. Such an internalization, by the young self, of the old self's interests could lead to the absence of a disagreement among selves regarding the old self's best action. But that case is only one case among many others. One cannot exclude *a priori* the possibility of disagreements among temporal selves of the same person.

[13] This is the case, for example, when $x = 2$, $y = 1$, $z = 0.50$ and $p = 0.75$. In that case, the young self commits suicide in game I, but remains alive in game II.

Whether or not allowing for extra flexibility for the old self encourages the young self to commit suicide depends on how the young self evaluates, *from his own point of view*, the reaction of the old self to the occurrence of bad news. If a disagreement between the two selves arises, flexibility for the old self can, in the absence of trust, encourage suicide of the young self. The reason lies in the fact that the old self cannot commit to do the particular action that is preferred by the young self. Hence, given the absence of commitment device and the absence of trust, the young self may decide to "close the game", in order to avoid that the old self deviates from the life plan that the young self regarded as the best.[14]

In sum, instead of regarding suicide as an irrational act of not waiting long enough before making an irreversible investment in non-existence, the theory of life as a trust game regards suicide as a rational act favored by the conjunction of four elements: (1) a disagreement between selves concerning the best action of the old self if bad news occurs; (2) a high probability of bad news; (3) the absence of a commitment device guaranteeing that the old self will respect the young self's plan; (4) a lack of trust of the young self in the old self.

4. Concluding remarks

Considering the decision to commit suicide (or not) as a trust game played by multiple selves of the same person casts original light on the suicide decision. Unlike what *The Option Value of Life* suggests, more future flexibility does not necessarily discourage suicide now. At this stage, it may be useful to turn back to a question raised in Section 1: how can it be rational to commit suicide now, given that remaining alive leaves open the possibility to commit suicide in the future in case of bad news? The trust theory of life provides the following answer: if the young self regards the suicide of the old self as a lack of loyalty towards him, he may, in the absence of commitment device, prefer to "close the game" now, because he does not trust the old self, and, hence, prefers not investing in the existence of the old self. In the absence of trust, a larger degree of flexibility in the future can encourage suicide right now, in contradiction with the theory of option value.

To conclude, it should be stressed that our theoretical result – the indeterminate effect of future flexibility on the suicide decision – has implications when thinking about the design of end-of-life policies. Whereas the theory of option value implies that guaranteeing access to euthanasia at very old ages encourages disabled individuals to remain alive, the theory of life as a trust game states that this may not necessarily be the case. This depends on how the disabled perceives the act of committing suicide at older ages. If that act is regarded as a lack of loyalty of the old self with respect to past investments made by younger selves in his existence, then guaranteeing future access to euthanasia may, in the absence of trust of the young self in the old self, encourage – rather than discourage – suicide right now.

[14] The nature of the disagreement among selves may take various forms. Instead of the example of Figure 1, one may, for instance, assume that the young self prefers that the old self commits suicide in case of bad news (i.e. $z > y$) and that the old self prefers not committing suicide in case of bad news ($b > c$). Under a high probability of bad news, the young self can decide to commit suicide, to avoid the old self from deviating from his ideal plan.

ACKNOWLEDGEMENTS

The author would like to thank Matthew Adler, Didde Boisen Andersen, Axel Gosseries, Simon Laumann Jorgensen, Kasper Lippert-Rasmussen, Soren Midtgaard, Shlomi Segall as well as two anonymous reviewers for their helpful comments and suggestions on previous versions of this paper. This paper also benefited from suggestions from participants of the *Northern Lights Workshop in Political Philosophy #3*.

REFERENCES

Berg, J., Dickhaut, J. and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10: 122-142.

Burri, S. 2020. The option value of life. *Economics and Philosophy*, forthcoming.

Diamond, P. and B. Köszegi. 2003. Quasi-hyperbolic discounting and retirement. *Journal of Public Economics* 87: 1839–1872.

Dixit, A. and R. Pindyck. 1994. *Investment under Uncertainty*. New-York: Princeton University Press.

Hershfield, H., Goldstein, D., Sharpe, W., Fox, J., Yeykelis, L., Carstensen, L., and J. Bailenson. 2011. Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research* 48: 23-37.

Laibson, D. 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112: 443–477.

Parfit, D. 1971. Personal identity. *Philosophical Review* 80: 3–27.