

## **Expectancy and rational action prior to personal fission**

**Paul Tappenden<sup>1</sup>      7 January 2010**

### ABSTRACT

Some analyses of personal fission suggest that an informed subject should expect to have a distinct experience of each outcome simultaneously. Is rational provision for the future possible in such unfamiliar circumstances? I argue that, with some qualification, the subject can reasonably act as if faced with alternative possible outcomes with precise probabilities rather than multiple actual outcomes.

In recent decades the idea of personal fission has been prominent in discussions of the identity of persons over time and imaginary scenarios have been presented in various ways such as divided brain hemisphere transplants and malfunctioning Star Trek teleporters. This theoretical work has spawned some analyses which suggest that a subject about to knowingly undergo fission should expect to have a distinct experience of each outcome simultaneously. If any of these analyses is to be taken seriously, or any novel view with the same implication, then it is an interesting question whether a subject in such a predicament could make rational future-directed decisions. A surprisingly simple argument shows that it is plausible that rational action is possible and would involve taking an attitude towards the expected actual outcomes which bears some similarity to normal probabilistic expectancy towards a corresponding range of possible outcomes with associated probabilities.

If any further motivation for tackling this question should be thought necessary it can be found in the context of the current lively debate about the Everett interpretation of quantum mechanics, often referred to as the many-worlds interpretation. My own concern with personal fission arises from that source and I shall make a couple of footnote references to papers in that literature but there will be no need here for any more extensive discussion of what is involved. Readers wishing to find out more on the connection with quantum mechanics might start with David Wallace's recent introduction (2008). The fission I shall be discussing might be called classical in the sense that it can be understood in terms of what is known as classical mechanics. There are similarities and differences between this and the quantum case.

My aim here is to present an idea, not to review the analyses of personal fission to which it might be applied. I shall be articulating the argument in the context of Ted Sider's 'stage theory' (1996, 2001) which I shall outline in a moment. I shall use that theory of trans-temporal identity because it is transparent and straightforwardly implies that a subject facing fission should expect to have each of the output experiences simultaneously as distinct people. Sider's theory is also suitable

---

<sup>1</sup> paulpagetappenden@gmail.com

because it has become fairly popular, though it is of course not without its critics. Other analyses of fission to which the following argument may apply can be found in Gallois (1998), Merricks (1997) and Perry (1972).

According to Sider's stage theory what a person is at any time is a momentary stage in their history. Stage theory embraces the idea of temporal parts but rejects conventional reductionism about identity which identifies a person's body with an aggregate of temporal stages, a so-called world-tube or spacetime worm. This feat is achieved by exploiting the idea of counterparts which was introduced by David Lewis in the context of modal logic (1968). Temporal stages in the past or future of a person or object are treated by Sider as a type of counterpart, so-called temporal counterparts. Thus I was a boy who scrumped apples because I have past temporal counterparts who scrumped apples. My having those temporal counterparts makes it true that I was that boy just as, given Lewisian possible worlds, my having modal counterparts with blond hair makes it true that I might have been blond.

According to stage theory it is true that I will be bald if and only if I have future counterparts who are bald. If I'm worried about being bald then why, an objector to stage theory might ask, should I worry about having a bald future counterpart? After all, that counterpart is not me, he is someone else. The stage theorist might respond that I should be worried about my bald future counterparts because they are people who I will be. I shall be assuming that some such response is satisfactory but I am not here claiming that stage theory is the best analysis of trans-temporal identity. I am using it as a vehicle for an argument which might be found applicable to theories other than Sider's.

For the sake of conceptual simplicity, imagine a case where a person's body is caused to divide smoothly in an amoeba-like way but with each fission product being isomorphic to the original, so having the same weight, volume and so on. This could be thought of as a complex process which involves the splitting of each constituent molecule into two molecules of just the same type followed by the spatial separation of the two sets of molecules. Such a process would require a huge input of energy, doubling the total corporeal mass involved, but fission of this type is not obviously physically impossible. The subject is anaesthetised and their body is put into a hi-tech splitter. The two emergent bodies are then awakened in rooms of different colours, one red and the other green.

We shall be wanting a hypothetical subject who, for whatever reason, favours the stage-theoretic analysis. The Tuscan sculptor Stagio Stagi had the right name to be a stage theorist if anyone ever had so let's call our hypothetical subject Stagio. In matters of trans-temporal identity Stagio embraces Sider's stage theory. According to stage theory, if Stagio is about to enter the splitter it is true that he will be the person who awakes in the red room and that he will be the person who awakes in the green room (Sider 1996. §II; 2001, p.201). That is because Stagio entering the splitter (a temporal stage) has a future counterpart in the red room and a future counterpart in the green room and Stagio prior to fission has the relation *will be* to each of those counterparts. The fact that each of those simultaneous future counterparts is a distinct person is not at odds with this result. According to stage theory then, Stagio prior to fission should expect to awake in the red room and expect to wake in the green room.

What is Stagio to make of this unusual prospect? For instance, if a philosophers' demon forces him into the splitter and convincingly tells him that the person in the red room will be held prisoner and tortured and the person in the green room will be free to leave how worried should Stagio be, believing that these are persons who he will become? Should he be more worried than if he were not facing the splitter but rather were told that he would be taken to the green room and set free? Should

he be less worried than if he were not facing the splitter but rather were told that he would be taken to the red room and tortured ?

A possible reaction here might be to suspend belief because the idea of expecting to have two incompatible experiences simultaneously as distinct persons is so strange that we lack the conceptual equipment to know what to make of it. But that could seem to be a strategy of despair and there is a way forward. To see how, call the person who awakes in the red room StagioR and the person who awakes in the green room StagioG. These are people who were both previously Stagio. And imagine that when StagioR and StagioG are awakened the lights are out in their rooms so that they are in complete darkness and Stagio knows in advance that this will be so. It would seem that under these circumstances Stagio should expect to experience waking in complete darkness since that is the common experience of StagioR and StagioG.

Now, StagioR in the dark can very plausibly assign probabilities as to whether he is in the red room or the green room<sup>2</sup>. Given the symmetry of the setup it seems reasonable that he should assign equal probabilities to each possibility and so a probability of a half to each. StagioG should do likewise. So StagioR and StagioG should each think he has an evens chance of finding himself in the red room when the lights are switched on.

In which case each should be worried to the extent they would be if the demon were to toss a fair coin to decide between torture and freedom. So, if Stagio knows in advance that the lights will be out for a while before the fates of StagioR and StagioG are decided it looks as though he should be confident that he will experience being just as worried about torture as if he were facing a decision based on the tossing of a fair coin rather than the fission setup. It appears that Stagio prior to fission has good reason to feel about his prospects in just the way he would feel if he were not facing fission but knew that the demon would decide his fate on an upcoming coin toss.

Note that the implication is not that Stagio, prior to fission, should assign a credence of one half to waking in the red room. We have already seen that according to stage theory he should fully believe that he will wake in the red room and that he will wake in the green room. Rather, the idea is that Stagio prior to fission, knowing that he will find himself temporarily in the dark, has good reason to view his prospects *as if* he were facing torture or freedom on the toss of a coin.

Some qualification is called for here. StagioR being tortured knows that StagioG is enjoying freedom and that might temper StagioR's suffering. Similarly, StagioG set free knows that StagioR is being tortured and so StagioG's relief could be spoiled by that thought. This is different from the coin toss situation. If torture or freedom is decided on a coin toss only one or the other of the options is actualised, no both, so the tortured person does not have any reason to think about a free one and *vice versa*. However, it is not obvious that this is enough to completely undermine Stagio's ability to adopt a rational attitude towards his fission when he knows that the lights will be out for a while. To see why we can consider how Stagio might be motivated to lay stakes if the red and green rooms were associated with a payoff regime like the outcomes of coin tosses on which bets are placed.

Before going into this we need to reflect on what becomes of Stagio's money when he enters the splitter. Being well informed, StagioR knows that there exists a StagioG who has an equal claim on his bank account and *vice versa*. So Stagio should expect in advance to face financial uncertainty since

---

<sup>2</sup> This idea is taken from Lev Vaidman (1998) who employs it in the context of the Everett interpretation of quantum mechanics.

that is common to the experience of StagioR and StagioG. In order that fission should not in itself have financial consequences for Stagio it would be necessary to provide independent bank accounts for StagioR and StagioG each containing the same funds as were available to Stagio prior to fission. Assume that Stagio has that assurance.

We saw that StagioR and StagioG with the lights out each has good reason to think that he has an evens chance of being in the green room. Suppose that they were betting men and each were told that for any stake laid there would be a payoff of more than two-to-one to the person in the green room and nothing to the person in the red room. If confident in the honesty of the setup both StagioR and StagioG should accept this as a good bet though the above qualification also applies since when StagioR loses he will know that StagioG has won and *vice versa* and that thought may affect how they judge what are sufficient odds to make the bet worthwhile. In the case of betting on the toss of a coin any odds over two-to-one can seem worth taking but in the fission case it is not clear that that is so. However, it is implausible that mutual empathy between StagioR and StagioG should completely overrule any parallel between the fission and coin toss scenarios. For suppose that the payoff in the green room is a billion-to-one with the stake simply being lost in the red room. It is hard to imagine that StagioG's glee at having won a fortune for a relatively trivial stake would be tempered very much by the thought that StagioR had lost a few pennies, or even by the thought that StagioR may feel very miffed at not having won the jackpot. Bear in mind that StagioR and StagioG are distinct persons who could just as well be on different planets. It's true that they share a common history but that has been and gone by payoff time. In general, gamblers who win a bet are not too bothered about the people who lose.

What this seems to show is that the qualification to do with mutual empathy in the fission case makes it differ from the chance analogue on a sliding scale. In the betting case the empathy factor may affect the marginal payoff regime which is acceptable but it may have less and less affect on a punter's behaviour the better the offered payoff regime is. I shall return to a more extreme case akin to torture and freedom in a moment but first of all more needs to be said about betting.

So far we have seen that StagioR and StagioG both have reason to lay a stake if the lights are out and the payoff regime is good enough. But does this have any bearing on how Stagio should act prior to fission? Well, suppose Stagio is told what the payoff regime will be and that there will only be a payoff in the green room if a stake is laid prior to fission and suppose that that payoff regime is good enough to motivate StagioR and StagioG to take the bet when the lights are out. In that case both StagioR and StagioG would regret that a stake had not been laid if Stagio had not done so. So Stagio can be confident that after fission and prior to the lights going on he will feel regret unless he has laid the stake that he knows StagioR and StagioG would judge appropriate, which seems to be a very good reason for Stagio to lay that stake in advance.

But what if there were no period of blackout after fission ? Imagine that Stagio is told that for there to be any payoff a stake must be laid before fission and the payoff regime will be such that if there were temporary blackout then StagioR and StagioG would want a stake to have been laid. But now Stagio knows that there will not be a period of blackout. Straight after fission StagioR will lose and StagioG will win. In that case Stagio knows that StagioG will be glad to have won and StagioR will be sad to have lost. Lloyd Humberstone (1980) has pointed out that in ordinary betting situations a punter has good reason to expect that regret is inevitable. Unless it were impossible to have staked

more, StagioG will regret not having laid a larger stake. And StagioR will regret having laid a stake at all.

If regret is inevitable then it seems reasonable that Stagio should aim to minimise the regret which he knows he will come to feel. Note that StagioG, having won, feels more regret the lower the stake and StagioR feels more regret the higher the stake so it looks as though there is scope for compromise. A reasonable compromise is for Stagio to lay exactly the stake which he knows StagioR and StagioG *would* judge appropriate *if* they were in the dark. Arguably a punter in a normal betting situation reasons in a similar way. The stake s/he judges appropriate to lay weighs the regret s/he would feel at not having staked more if winning against the regret s/he would feel at not having staked less if losing. That being so, it looks as though Stagio is motivated to lay a stake in advance if the payoff regime is good enough even if there is no prospect of being in the dark after fission. The very possibility of being in the dark after fission is enough to guide Stagio's reasoning beforehand.

In order to be clear about what I'm claiming here it may be useful to consider a hybrid case involving both chance and fission. Suppose that Stagio is told that a fair die will be rolled to determine whether he undergoes fission or not. If the die comes up 6 he will not undergo fission and otherwise he will go into the splitter as before. In this case Stagio, prior to the roll of the die, judges that the probability of becoming StagioR is  $5/6$  and the probability of becoming StagioG is  $5/6$ . The fact that these probabilities sum to  $5/3$  does not reveal an inconsistency but just the strangeness of stage theory as applied to fission. Recall that when fission is certain Stagio prior to fission judges that the probability of becoming StagioR is 1 and the probability of becoming StagioG is 1.

If fission takes place after the roll of the die but StagioR and StagioG are in the dark and do not know whether or not fission has indeed occurred then each will judge the probability that fission has occurred as  $5/6$  and so, given the symmetry of the setup, that the probability of being in the red room is  $5/12$ . That is the probability that StagioR and StagioG will take into account if offered a bet and so what Stagio needs to take into account if the argument which I have presented is good. But note again that the implication is not that Stagio prior to the die roll should expect, with a credence of  $5/12$ , to awake in the red room. At best the argument shows that Stagio should lay a stake *as if* the chance of awaking in the red room were  $5/12$ . To repeat, according to stage theory, prior to the roll of the die, Stagio should expect with a credence of  $5/6$  to awake in the red room and should expect with a credence of  $5/6$  to awake in the green room. Note also that he should assign a credence of zero to waking in the red room *and* the green room according to stage theory since no one person awakes in the two rooms simultaneously.

This last point deserves a little more discussion as it is *prima facie* very strange. The problem is that it can be tempting, putting oneself in Stagio's place, to try to imagine what it would be like to expect to awake in a red room and in a green room simultaneously and to mentally grope for an intuition which could help guide one's judgement about how best to act. That's rather like a novice aircraft pilot who is tempted to think that it is possible to fly in cloud without instruments. But it is just not possible; without a gyroscopically driven artificial horizon the pilot is bound to lose control because the sensibilities developed on *terra firma* play tricks. The more firmer the less terror, as pilots say. What I'm claiming is that, for Stagio, imagining the judgements StagioR and StagioG would make if they were in the dark acts as a sort of mental artificial horizon which can steer decision-making though fission.

Sider's theory suggests that a person should assign a credence of  $X$  to having a given future experience if and only if the probability of having a temporal counterpart having that experience is  $X$ . That implies that Stagio, facing fission with probability  $X$ , should assign a credence of  $X$  to awaking in the red room and a credence of  $X$  to awaking in the green room. But on Sider's theory 'I will awake in the red room and I will awake in the green room' is not logically equivalent to 'I will awake in the red room and the green room' on our usual understanding because the latter phrase is normally taken to imply that one person will be in both rooms. According to stage theory the persons in the two rooms are distinct albeit having a common origin. The credences which Stagio, prior to fission, must assign to future experience simply do not add up to being credences which he can use to guide his future-directed action. They are logically coherent within the framework of stage theory but for all practical purposes they are useless, rather like a pilot's sense of orientation in cloud. Stagio needs to keep his mind's eye on the artificial horizon, the credence assignments StagioR and StagioG would make if subject to the appropriate ignorance conditions. It is those credence assignments which can help to guide Stagio's decision-making prior to fission if the arguments which I have presented are correct.

We have been imagining fission as a symmetric bifurcation of Stagio's body but splitting into three, four or even a denumerable infinity of output bodies ought to be allowable if bifurcation is. If there were a number  $N$  of distinct outcomes then output persons held in a state of ignorance about their outcome should very plausibly assign a probability of  $1/N$  to that outcome being a particular one of the range. Further, if there were  $N$  outcomes where a proportion  $A/N$  were outcomes of type A, a proportion  $B/N$  were outcomes of type B and so on then the post-fission, pre-observation subject should assign a probability of  $A/N$  to seeing an outcome of type A and  $B/N$  to seeing an outcome of type B etc..

This thought can be used to demonstrate another way in which Stagio's reasoning facing fission cannot be exactly as it would be in facing corresponding chancy scenarios despite the similarities which I have discussed so far. Imagine that Stagio had lost a leg to a shark and a special new fission treatment were available which had an unavoidable defect<sup>3</sup>. The treatment involves a thousandfold fission where 999 output persons have two good legs but one is rendered legless. If Stagio were facing a treatment which involved a one-in-a-thousand chance of leglessness he might well take the risk. In the fission case, if the patients were held in a temporary state of ignorance they would each suppose that the chance of being legless was one-in-a-thousand but Stagio would know in advance that there must be a legless person. Could he go ahead with the treatment in the fission case in the same way that he might in the corresponding chancy case? It's not clear he could. Moral issues arise which are absent in the chancy case. In opting for the fission treatment Stagio can be sure that someone will find himself legless, though perhaps that thought is mitigated to some extent by the fact that Stagio would believe in advance that if he went ahead with the treatment the legless person would be aware that he himself was the person who signed the consent form.

That thought experiment was based on the idea of fission into many outcomes. The idea of multiple splitting introduces a further complication. Imagine that Stagio is faced with a three-way split with StagioR and StagioG arriving in red and green rooms as before but also with a StagioB in a blue room. If the lights were out StagioR, StagioG and StagioB would each put the odds at one third of

finding himself in the red room. But now suppose that Stagio were told that rather than a straight three-way split there would be a double two-way split, all taking place on anaesthetised subjects. First of all Stagio would fission into StagioR, who would be wheeled straight into the red room, and StagioX who would then go on to fission into StagioG and StagioB. In that case, whilst the lights were out, it would seem compelling that StagioR, StagioG and StagioB should put the odds at one half of being in the red room, one quarter of being in the green room and one quarter of being in the blue room. So if Stagio is faced with a scenario where he needs to factor in the possible credence assignments of post-fission, pre-observation persons who he will become he needs to know exactly what the form of the fission setup is in order to be able to do that.

Finally, there is another class of possible fission scenarios which needs to be taken into account. Suppose that Stagio is told that in the red room, rather than being awakened, he will be instantly and painlessly killed. In that case the post-fission subject in the unlit room can be certain that when the light is switched on he will see green. And Stagio pre-fission can be sure that he will survive the process and see green. That is not at all like the equivalent chance setup where a person is faced with a fifty-fifty chance of dying. What this points up is the fact that the post-fission subject who knows fission has occurred but remains ignorant of the outcome knows that he is not dead so his judgement of the probabilities of seeing various outcomes must be 'renormalised' to take this into account. That is, in a case of one-stage fission into N outcome branches where some number D of them involve instant death of the subject then the full range of branches for which the post-fission subjects in a state of ignorance about outcomes must use as a basis for calculating proportions is not N but rather N minus D.

The upshot of all this seems to be that Stagio's predicament facing fission is not clear-cut but he need not necessarily be excluded from making rational provision for the future either. It looks as though there is a body of cases where Stagio might reasonably act much as if he were facing a range of possible outcomes with associated probabilities rather than a range of actual outcomes but the idea does not apply to all imaginable scenarios.<sup>4</sup>

## References

- Gallois, A. (1998). *Occasions of Identity*. Oxford, Clarendon Press.
- Humberstone, I. L. (1980). You'll Regret It. *Analysis*, 40, pp. 175-6.
- Lewis, D. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy*, 67, pp. 427-46.
- Merricks, T. (1997). Fission and Personal Identity Over Time. *Philosophical Studies*, 88, pp. 163-86.
- Perry, J. (1972). Can the self divide? *Journal of Philosophy*, 69, pp. 463-88.
- Price, H. (2010). Decisions, Decisions, Decisions: Can Savage Salvage Everettian Probability? In *Many worlds: the Everett interpretation of quantum mechanics*. Saunders, S., Barrett, J., Kent, D., Wallace, D. (eds.), Oxford, Oxford University Press. Available online at <http://philsci-archive.pitt.edu/archive/00003886>

---

<sup>3</sup> This idea comes from Huw Price (2010)

- Sider, T. (1996). All the world's a stage. *Australasian Journal of Philosophy*, 74, 433-53.
- Sider, T. (2001). *Four Dimensionalism*. Oxford: Oxford University Press.
- Vaidman, L. (1998). On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory, *International Studies in the Philosophy of Science*, 12, 245-66. Available online at : <http://xxx.lanl.gov/abs/quant-ph/9609006>
- Wallace, D. (2008). The Quantum Measurement Problem: State of Play. In *The Ashgate Companion to Contemporary Philosophy of Physics*, Dean Rickles (ed.) London, Ashgate. Available online at <http://philsci-archive.pitt.edu/archive/00003420>

---

<sup>4</sup> My thanks to Lloyd Humberstone and particularly to an anonymous referee for very useful comments on previous drafts.