On the History and Philosophy of the Black Hole Information Paradox

Sebastian De Haro and Jeroen van Dongen
University of Cambridge and University of Amsterdam

Seven Pines Symposium, Minnesota, 18 May 2017

Preliminary Remarks

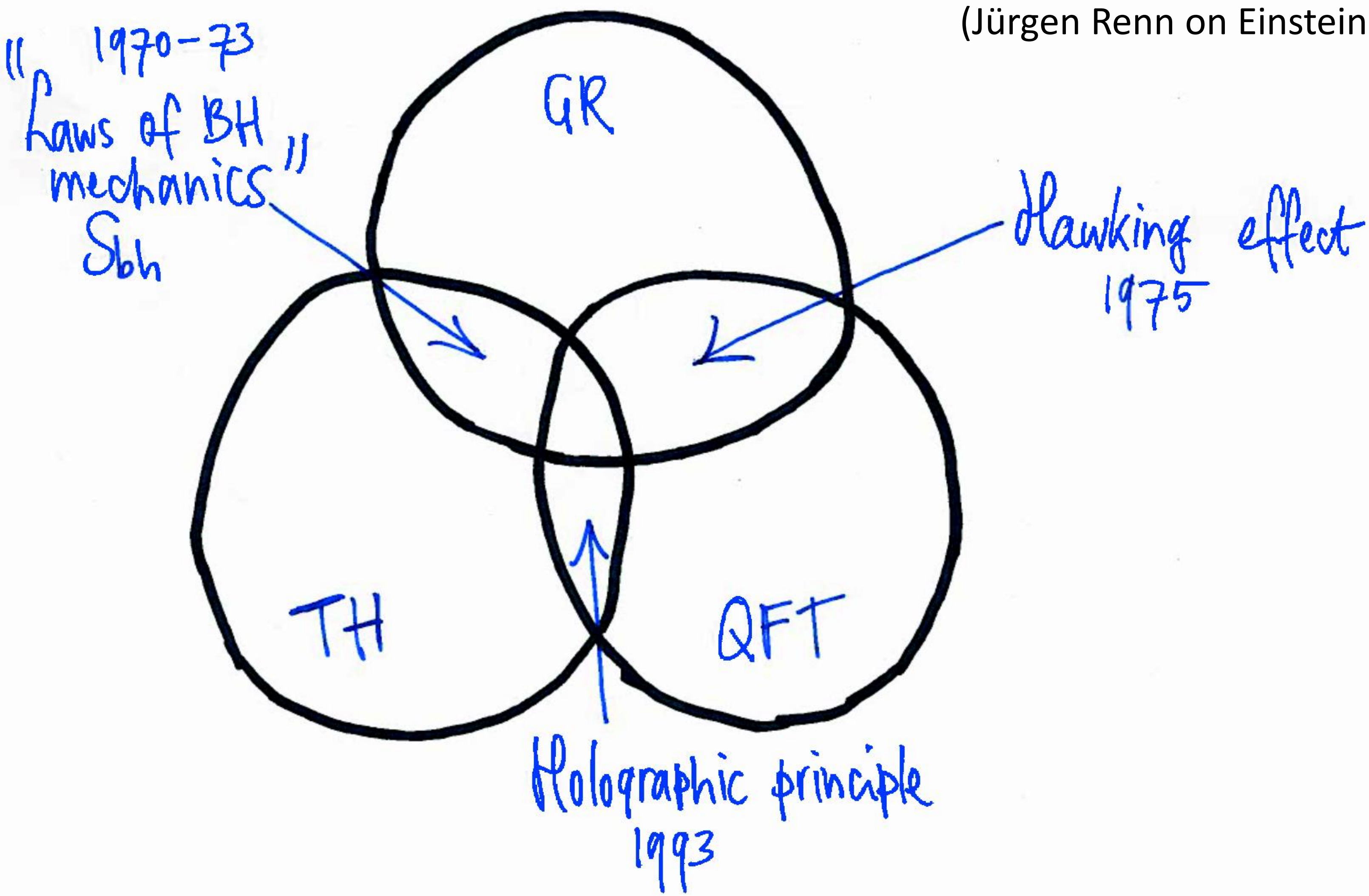
- Historical (HPS) perspective of the talk:
  - Detached view about how the debate went.
  - Aim: to understand the *dynamics of the debate,* rather than assessing who is *right*.
  - A case-study of the dynamics of a debate that is of scientific importance, involving two significant physics communities: relativists-quantum field theorists, and particle physicists.
  - Thus the absence of a concrete solution to the information paradox (and the absence of a consensus about what a solution should look like) is unimportant.
- I will follow the historical sequence of the events.

Questions

- Why were black holes central objects of theoretical study?
- How did theorists talk about black holes and converse with each other?
- Can that inform us about the *nature* of the information loss debate?
- Can that inform us about rationality in the context of theory development and assessment? What does that tell us about *how science works?*

The black hole as a problem at the border

(Jürgen Renn on Einstein 1905)



Early Years, 1968-1981

Pre-history

- Early results on extraction of energy from rotating black holes: Penrose process (Penrose, Christodoulou: ’68-’70).
- Early results on QFT in classical curved spacetimes (in particular, cosmological models): Parker, Fulling, Unruh (‘69-’74).
- The two communities were disconnected, despite their institutional connections, in ‘71-’72 (Fulling under Wightman at Princeton; Unruh, Christodoulou, and Bekenstein under Wheeler):
  - Different ways of looking at the black hole: as dynamic (formation, singularity, horizon) or as background where quantum fields and the stress-energy tensor are evaluated.
  - No cross-citations during this period.
- Laws of black hole mechanics (Bardeen, Carter, Hawking, ’70-73):

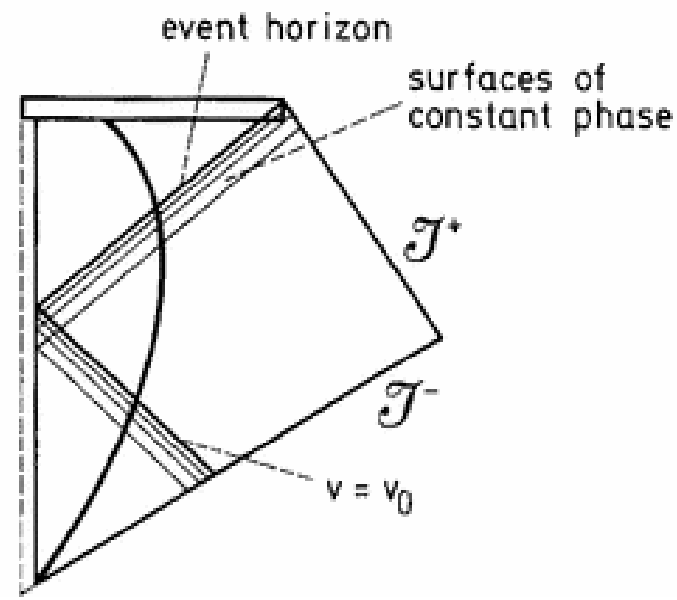
$$\mathrm{d}M = \frac{1}{8}\kappa \mathrm{d}A + \Omega_H \mathrm{d}J \qquad \text{and} \qquad \delta A \geq 0$$

  - Disanalogies: not in thermal equilibrium, zero temperature.
- Bekenstein (’72-73) Generalised Second law: $S = \frac{kc^3 A}{4G\hbar}$

Hawking radiation ('74-75)

- At $\mathcal{J}^-$: $\phi = \sum(fa + \bar{f}a^+)$ and At $\mathcal{J}^+$: $\phi = \sum(pb + \bar{p}b^+)$

  Then: $a = \sum(\alpha b + \beta b^+)$

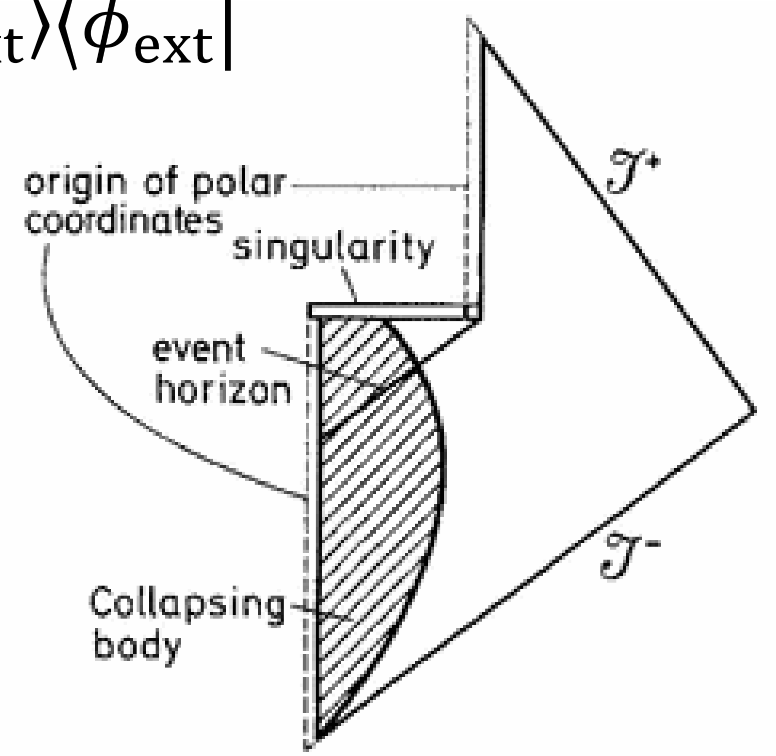


Early developments on Hawking Radiation

- Wald 1975: exact black body thermal state.
  - Rigorous quantum field theory methods.
  - Density matrix identical to black body spectrum.
  - Result might shed light on a deep connection between entropy and black hole surface area.
- Unruh 1976: particle detection, acceleration.
  - Renormalized stress-energy tensor in black hole metric.
- Hawking radiation is an established result.
- Flourishing of quantum field theory in curved spacetime.

Hawking 1976: Information loss

- PRD 1976: "Breakdown of predictability in gravitational collapse".
- $\psi \in \mathcal{H}_{\text{int}} \otimes \mathcal{H}_{\text{ext}}$
- $\rho_{\text{ext}} = \text{Tr}_{\mathcal{H}_{\text{int}}} |\psi\rangle\langle\psi| = \sum p \big|\phi_{\text{ext}}\big\rangle\big\langle\phi_{\text{ext}}\big|$
- Pure → Mixed state.



Early responses (pre-1982: interviews 2012) to Hawking's 1976 Unpredictability

- Wald: "When I heard Hawking's info loss result, I thought, 'Wow! Neat!' Quickly accepted by the dozen or so GR people with expertise in this area."
- 't Hooft: "Created no waves among particle physicists. I did not know about it."
- Unruh: "The information paradox was not a meteorite hitting the earth. Rather: a bunch of participants developing a story."
- Citation analysis: QFT & GR communities.
  - Ignored or unknown by particle physicists.

Citations Hawking's 1976 before 1985: 67 records

1. **Breakdown of Predictability in Gravitational Collapse**
S.W. Hawking (Cambridge U. & Caltech). 1976. 14 pp.
Published in **Phys.Rev. D14 (1976) 2460-2473**
DOI: 10.1103/PhysRevD.14.2460
References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndN
ADS Abstract Service; OSTI Information Bridge Server
Detailed record - Cited by 1314 records 1000+

refersto:recid:114952 and date before 1985
find j "Phys.Rev.Lett.,105*" :: more

Sort by: earliest date | desc. | - or rank by -
Display results: 25 results | single list

HEP **67** records found 1 - 25 jump to record: 1

1. **Nontrivial Topologies in Quantum Gravity**
S.W. Hawking (Cambridge U.). Aug 1984. 12 pp.
Published in **Nucl.Phys. B244 (1984) 135-146**

By 1990: 105

refersto:recid:114952 and date before 1990 | Brief format | Search | Easy Search | Advanced Search
find j "Phys.Rev.Lett.,105*" :: more

Sort by: earliest date | desc. | - or rank by -
Display results: 25 results | single list

HEP **105** records found 1 - 25 jump to record: 1

1. **On The Possibility Of Detecting Wormholes In A Quantum Coherence Experiment With A Rf Squid**
Subhendra Mohanty (Wisconsin U., Madison). 1989. 57 pp.
RX-1270 (WISCONSIN), UMI-89-24087

By 1996: 284

refersto:recid:114952 and date before 1996 | Brief format | Search | Easy Search | Advanced Search
find j "Phys.Rev.Lett.,105*" :: more

Sort by: earliest date | desc. | - or rank by -
Display results: 25 results | single list

HEP **284** records found 1 - 25 jump to record: 1

1. **A Noncritical string (Liouville) approach to brain microtubules: State vector reduction, memory coding and ca**
N.E. Mavromatos (Oxford U.), Dimitri V. Nanopoulos (Texas A-M & HARC, Woodlands). Dec 1995. 70 pp.

Late 70s QG: disparate fields

- Meetings across fields? Discussions?
- Unruh: “No. Particle physicists were **not interested in gravity** from the 1930s through the 1970s.”
- ’t Hooft (2012): “**Quantum gravity struck me as highly esoteric.** No observations; what on earth are all those people doing?”
  - “Relativists protested the way I treated [the metric in renormalization study]. **The metric was absolutely sacrosanct!**”

1982-1989

Particle Physicists

Early 1980s: arrival of particle physicists

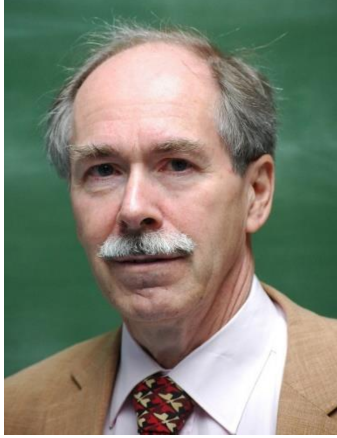
- Susskind (2008): 1981 EST meeting San Francisco.
- ’t Hooft unsure: earlier history 1970s conferences.
- Unruh: “Suddenly these particle physicists turned up and

there was a problem. I still don’t get what was so problematic.”

Black Hole Conservation Laws

- Ellis, Hagelin, Nanopoulos, Srednicki (1983): Modified Hamiltonian equation for density matrices: used to interpret the bounds on the violations of quantum mechanics set by the then available experimental data.
  - In Hawking's scenario, "symmetry rules do not necessarily lead to conservation laws". In particular: energy non-conservation.
- Used as an argument against information loss: Gross (1984), Banks, Susskind, Peskin (1984).

’t Hooft’s papers



- 1985 onwards, ‘*S* matrix Ansatz’:
  - “We start with the **postulate** that there exists an extension of **Hilbert space comprising black holes**, and that a Hamiltonian can be precisely defined in this Hilbert space.”
  - Investigate the **S matrix**, by focusing on interactions between infalling matter and outgoing radiation: principle QFT approach.
  - Cf. Polchinski (2015).

The black hole as a borderline problem: ’t Hooft 1985

- “In 1985 I wrote: here is a **paradox.** And paradoxes are important! While searching for repairs, you can make new physics.” [Interview 2012]
- Like Einstein: from a *constructive* to a *principle* approach.

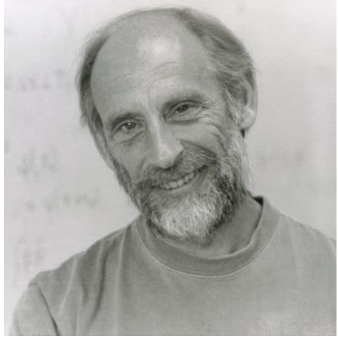
Principle approach

1993: Holography (new principle found):

“Given any closed surface, we can represent all that happens inside it by **degrees of freedom on this surface** itself.”

“This, one may argue, suggests that quantum gravity should be described entirely by a **topological quantum field theory**, in which all physical degrees of freedom can be projected onto the boundary.”

1980s: a clash of principles without a broad debate



- "Stephen Hawking had put his finger on a ***clash of principles***. The Equivalence Principle and Quantum Mechanics were on a collision course."
- Susskind: Particle physicists uninterested. "Complacency bothered me. This was the great problem of our generation!"
- 't Hooft: "For a while I felt that **I stood almost alone. Particle and string theorists were not interested.** Thought the black hole was some kind of soliton. None of their business, they thought."

1980s: still poor communication

- ’t Hooft & Unruh: “**little communication between communities**”.
- Wald: “Spoken with **maybe 6 particle physicists**. In [2010] I finally spoke with Banks at **Seven Pines**” (interview, 2012).
  - “Different in ‘firewall’ debate!” (Wald 2016).
- Unruh on ’t Hooft: “could not follow those articles at all.”
- Unruh on BSP: “Simply wrong; complete red herring. Black hole decoherence does not imply energy non-conservation.”

Early 1990s

String Theory

Early 90s string theory interest: first new calculations

- Callan Giddings Harvey Strominger (1992): ‘evanescent’ black hole 1+1 dimensions.
- Bañados Teitelboim Zanelli (1992): black hole in 2+1 dimensional AdS.
    - Spawned interest in the information paradox in the string theory community.
- Unruh: New interest in info paradox? “I did not follow that literature.”
- ’t Hooft: “When string theorists finally got started [in 3 + 1 dim], **they got it completely wrong.** Remnants!”

Key moments in 1990-1994

- 1990 “on the radar” and few extra QFT people on ’t Hooft-Susskind side.
- 1993 Black holes are a focal point of research due to CGHS black holes that claim to solve the information loss problem.
- 1993 Black Hole Complementarity: Susskind, Thorlacius, Uglum (proposal relates to ’t Hooft): principle approach, but also (1+1)-dimensional dilaton model.
  - Wald: “Violates local laws of QM. Radical idea to solve a problem I don’t see as radical”.
  - Susskind: “Info paradox arrived in a big way”; poll numbers start shifting, reach break even point.
- 1993-4: Still not many calculations one can do; but also holography.

After the second string revolution, 1995

Key moments 1995-2003

- 1995 Second String Revolution: M theory and D-branes.
- 1996 Vafa-Strominger calculate extremal black hole entropy (in a unitary theory).
- 1996 Callan-Maldacena calculation for near-extremal black holes convinces many senior string theorists.
- 1997 Maldacena AdS/CFT.
- 1998 Witten: AdS/CFT is holographic.
  - Maldacena argues that the black hole is unitary at the boundary.
- **Communication lines change.** Relativists also work on string theory. Joint conferences.

Susskind's language

- <u>Military</u> (*The Black Hole War and My Battle with Steven Hawking to Make the World Safe for Quantum Mechanics*); "neutrals" turned "allies"; and <u>Emotional</u>.
- <u>High stakes:</u> "Clash of principles", "Holographic Principle".
- <u>Certain:</u> "[Maldacena and Witten] proved that beyond any shadow of a doubt that information would never be lost".
- <u>But NO deductive certainty:</u> status AdS/CFT and dictionary, AdS spaces, idealizations and approximations, number of dimensions, etc.

The role of training

- Susskind: “victims of our faith based illusions? […] It all came down to: which principles do you trust?” “Hawking was **too classically wired**”.
- ’t Hooft: “Hawking works rather abstractly. Euclidean gravity etc. I like **concrete things: particles**. Wald, too, is much attached to axiomatic QFT in curved spaces. He really has the relativist’s vision; of people that grew up in GR, have been **pampered by GR**” (2012 interview).
- Who has the proper *understanding,* and hence authority, about this problem? → Frustration.

The role of training

- Wald: “Most particle physicists are not used to non-Cauchy type evolution laws because they have always worked in flat spacetime. They do not start with a spacetime point of view, but it is awfully difficult to understand a black hole if you do not have a spacetime point of view.”
- Unruh: “Particle physicists’ training is strongly rooted in flat spacetime (no singularities, or issues with causality). Unitarity was hammered into them by their professors, so that they stopped thinking about it. Their thought processes are really all stuck down there in flat spacetime.”

AdS/CFT?

- Susskind: Satisfied. “The Holographic Principle” is not speculative anymore, but tool.
- Unruh (2012) on AdS/CFT: “I distrust the argument. Can I point to anything? No. I don’t understand string theory well enough.”
  - Cf. Unruh and Wald (2017).
- Wald (2012) on AdS/CFT: “It is completely unsatisfactory with regards to providing an explanation as to how things work locally.”

Today

- Hawking 2004: conversion.
  - Retrained himself in string theory.
  - Unruh (2012): “I was annoyed. Hand-waving arguments, following Maldacena, meagre results.”
- Information loss: minority of hold-outs.
- Yet: firewall debate since 2012!
- Hawking, Perry and Strominger (2016): hair at spatial infinity.

Understanding the debate

Paradigm shift?

- Borderline problem produces Anomaly*: the Paradox.*
- Holography and demotion of spacetime.
    - Incommensurability and Kuhn loss: not an explicit answer to the questions; requires rejection of QFT in curved space.
- More field theorists pour into the subject; tipping point.
- Paradigm shift but no clear *Gestalt switches*
    - Holography in the case of Hawking?
- Continuities too: semiclassical calculation and techniques still applied.
    - Allows for hold-outs, hybridity.
- Controversy as regress: cannot understand, don’t have the means, no resolution of the debates.

A debate over principles

- Recall the inductive gap between the evidence and the scientific claims: science as underdetermined by the available evidence (Duhem, Quine).
  - Kuhn (1977): no algorithm or set of rules for theory choice. The process of judgment regarding theories is better characterized as a **value judgment:** the characteristics desired of good scientific theories. By establishing a shared set of values used to inform theory choice, science can remain objective even if *some* values are essential to science (Douglas 2015).
- Black holes: *no experiment* and no deductive certainty: theorists' regress. The space between tested physics and theory to be filled by principles.
- Judging theoretical virtues as values:
- **Rationality** as weighing of principles in a gradual process.

- Not just ‘clash of principles’, but of ways of doing and being as well:
  - General relativity’s and AQFT’s rigour vs. speculative physics, heuristic arguments.
  - Treating the metric as ‘sacrosanct’ (‘t Hooft) or ‘being stuck in flat spacetime’ (Unruh).
- Conflict of *values*; **rationality as the weighing of values:** as a necessary element, further to mere prediction and testing.

The Firewall

- *Nature* (2013): "Another option, so controversial, that few dare to champion it: maybe Hawking was right all those years ago and information is lost".
- AdS/CFT:
  - Taken to support unitarity i.e. deny information loss.
  - Options left: give up low-energy effective field theory, or firewall.
  - Bousso: "Nobody wants to entertain the possibility that Maldacena is wrong".
  - Polchinski: "deepest ever insight into gravity"; like Maxwell's unfications.

A Philosophers’ Approach to the Debate

- Belot-Earman-Ruetsche (1999) argue that axiomatic QFT in curved spacetimes is fully consistent: hence, pro information loss.
  - Acknowledge that the debate has methodological catalysts, but are not interested in them.
- BER’s wording is careful, but their analysis does not explain the rationality of the debate (the analysis is at risk of rendering large parts of the debate as **irrational**).
  - Cf. also Maudlin’s recent preprint.

Conclusion

- Due to the novel, heavy role of “non-empirical theory assessment”, we need to “alter the philosophical understanding of the relation between a physical theory and the world” (Dawid 2013).
  - Not what we see in the information paradox, if what we try to understand is the *rationality of the debate*: rather than the truth about information loss!
- The Kuhnian picture of science applies well: rational science as a weighing of values (necessary element, additional to prediction and testing).

Thank you!