The Faces Of RS/RL: Looking into the liquid mirror of the archive

1 Introduction

This paper takes on a speculative and multilimbed approach to feel out the current figurations of recommender systems and their intelligent agent embodiments, in particular reinforcement learning. These technologies are vastly complex and are still surrounded by very conflicting opinions about how they operate and execute, the ethics and matterings they are able to be subjectively, as well as their purposes and functionality. By describing each of these faces of RS/RL I aim to weave between them a new envisaging of how these bodies might operate and be figured, moving past their complex and problematic histories and contemporary embodiments to imagine what practicing more collaborative, equal and open modes with them might look like.

2 Stirring the Pot

In Femke Snelting's essay "A fish can't judge the water"(2006), she constructs an analogy of technology as a wooden spoon to form an understanding of how technologies work as extensions of ourselves into phenomena. With the spoon you can enter into spaces not able by the body, and explore into liquids/layers not visible to the eye, "to feel at which moment exactly the starch starts to burn to the bottom of the pan"(Snelting,2006), but with this comes a change in ability, you cannot taste the flavour or feel the temperature of the sauce through the sauce, but you can with it.

"But do we use software to think?" (Snelting, 2006)

She expands the short essay into the political infrastructures of production and distribution of software, as well as their perceptual frameworks of standardisation and ownership. In this she explores how many levels of these apparatuses curtail our abilities within them, feeding our creativity through set uses and into predetermined outcomes. Countering this Snelting enjoys being immersed in a "a stockpile of milieus"(snelting, 2006), engulfed in exchange and emanating collaboration, "A milieu which supports biodiversity; a rich mixture of programmes and approaches".

When applying Snelting's analogy of the spoon to recommender systems (RS) it works two-fold, as these are the tools we use to stir the pot of archives/databases, they move through complex fluid and dynamic bodies, helping us to feel what is there, opening up these once static structures to new dynamic forms of organization. The second fold takes her approach to understand infrastructures and applies it to RS, examining how they makes us able to navigate archives, our worlds, and the memories/narratives we can create from them. Through this approach you see that RS's predominant development and financing by advertising corporates limits agents to forge paths to our most suited advertisement, instead of our own legitimate interests. Over iterations of use, they get to know us better, but not to explore or develop new parts of ourselves and our world but to reinforce set roles or encodings (Hall,) to make us better receptors, consumers and data suppliers. With this, we questions the RS spoon we have been given and asks how does it feel to stir with it? What does it make us able to sense and do? And how might we do it if we could carve it ourselves?

At the heart of recommender systems is the automation of choice, creating decisions about what to show and how. These operations are normally taken on by intelligent algorithms, from sophisticated hand calibrated decision models, through to AI models such as the focus of this essay,

Reinforcement Learning (RL), as well as other types of intelligent agents[1]. Practising through these technologies this research finds it essential to grab the spoon and feel the sauce to understand how technologies like RL frame their problem and model their matter. This essay aims to explore seven different faces that these apparatuses can be understood through, each of them contributing to a milieu of imaginaries that RL is and could be understood through. In this motion, I also intend to queery their perceptions and criptique their aims to form a perception of them that challenges us into new relationships and imaginaries of a respectful, collaborative, and sustainable RS/RL.

In this paper I specifically engage with RL models which are used in several different cloud based applications, across dynamics, play and recommendation/organisation systems. The work that I draw on in this paper is discursive of Facebooks/Metas (Gauci et al., n.d.), Googles (Covington et al., 2016; Cheng et al., 2016) and Amazons (Linden et al., 2003; Rybakov et al., 2018) many developments and projects in this area. These referenced papers give a foundational understanding of the status quo of RS/RL as a place of encoding taxonomies, and defining hierarchies, with these models automating and maximising growth through extraction, of resources, time and ability. This paper doesn't have enough space to fully trace RL's histories but through each of the papers sections I bring focus to specific elements of RL's figuration, reading into its construction, logics and implementations that operate through domination, regulation and punishment, aiming to bring about new ways of working that subvert these traditions.

2.1 framing/environment (what is this place where does it come from)

Many agent type mechanisms in computer science, RL Included, are framed within what's termed as an environment, this is the world/phenomena/problem that they have computationally modelled and must work through. In RL this can be theoretically understood as the dynamics that the agent needs to navigate to win/succeed as much as possible, learning a relationship of cause and effect formed within the agents' neural nets (NN). This environment could be an agent trying to get a ball in a goal for a game or for a recommender system, it would be more likely to be selecting one or few options from a database/generator to be shown to the user. Much like many other modelling problems within AI, ML and greater statistics, the world is itself often twisted and abstracted to fit into the materiality of computationally performed apparatuses and their formulaic ways of seeing/doing things. This is especially true for RL agents as the mathematically and volumetrically constructed problems have to be understood/reduced to the simplification of a (normally one dimensional) win lose values system (much like money in capitalism), meaning that complexities and subtleties within a problem can often be lost within the abstract inference of the environments vague scoring system, normally a single float (decimal number).

This embodying of environment as a problem originates from the punitive and carceral histories of RL, coming from Edward Thorndike's Animal Learning and the Law of Effect, which aims to reinforces behaviours and knowledge through penalisation and meritocracy. Starting off from animal learning where he trapped cats in boxes, with a task/problem for them to navigate and resolve that would lead them to escape the confines of the box and get a reward of sustenance. Thorndike then later evolved this into the law of effect which is a more complex version for discussing child learning, but still along the same punitive dialogue, reinforcing roles and understanding through reward and punishment. The essential problem that has stayed the same across all reinforcement mechanics, whether it be with animal, human or machine is that they are not only unnecessarily cruel, but they are sub-dimensional in their response to the outcomes to the agents actions, they point the finger,

[1] Genetic models, cellular models, Markov chain

but don't guide the hand. Punishment and reward in these terms will never be able to explain competently why what you or an agent did was wrong or how to improve, all that these low dimensional metrics and abstractions are able tell you is if you are doing good or bad in relation to other experiences in that set environment. This in some configurations takes different embodiments from more basic run/episode comparison to more complex and granular step comparison/ forecasting in SAC models, but their ability to navigate entropy and the unknown lacks through this penal apparatus that is only able to reinforce the know and standard.

With this in mind RL mechanisms are often only good at learning roles that can be prescribed through these win loose apparatuses, even still their mathematical environments/problems have to be carefully orchestrated to get the correct behaviour, meaning that the agents pick up the desired traits but also don't cheat or find a flaw within the environment's punitive logic. What if we could use a much more dynamic value systems as guides to help new knowledges evolve, ones that don't impose static and flawed idealisms, but instead construct environments that enable fruition and growth.

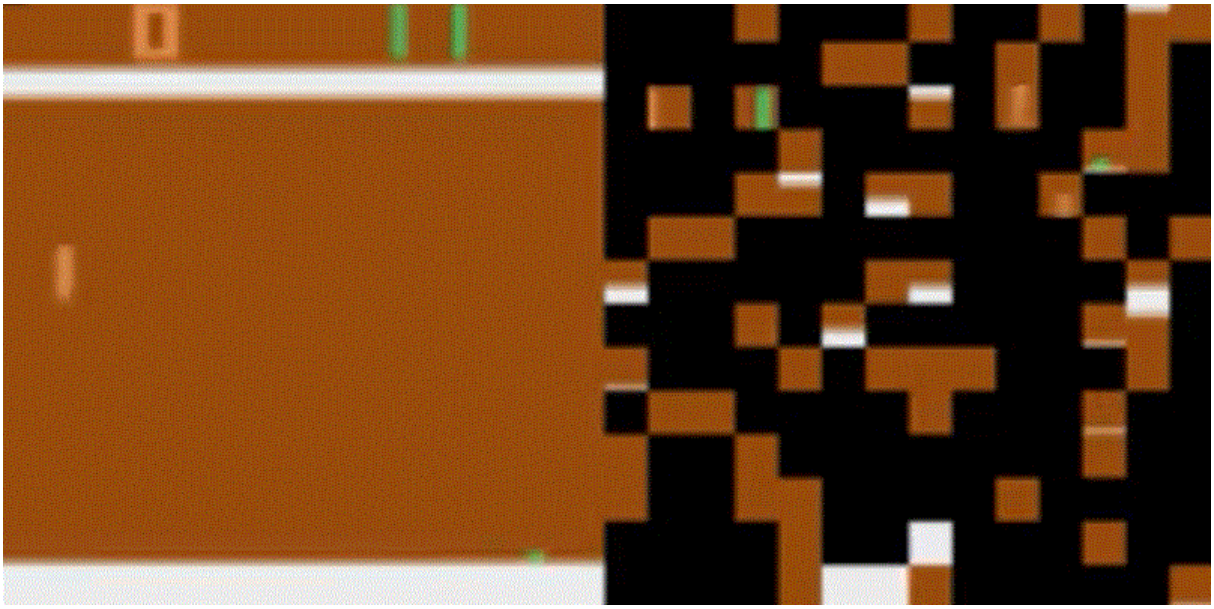
2.2 Agents (experiences and abilities of the agents in this space)

To explore the agents position and embodiment within RL It is interesting to take a position of Ian Bogost's Alien Phenomenology to fabulate the RL Agents experiences and interactions with its environment. The materially of this experience is through its inputs, outputs and a reward system. The inputs are the received data from the environment, normally an array of numeric data representing anything in the environment from the camera or vision of the agent in the environment to the last month of stock trading data. The outputs, otherwise termed actions are the numeric decisions the agent makes from passing the data through its NN, these actions in multi-step environment can have direct effect on its emergent future inputs and abilities. Like many NN these input and outputs are almost always abstracted and/or normalized variables transformed into either discrete variables (integer categories 1,2,3) = left, right, jump or continuous (floats -1…1) = the magnitude of something. These can be applied in any means of ways but must always maintain a clear relationship/problem for the third element of the penal reward system to sediment and converge the NN upon.

For instance if we take an RL used in recommender systems like YouTube or TikTok, we can imagine (with the help of papers) that they see/feel/have inputs that touch and sense your past selections, age, gender and geographic information, as well as the video put up for selection, with more complex inputs/formats being judged as a tokenistic numeric taxonomy (encoding) of the desired input transformed through an auto encoder. The Agent or model within this relationship outputs an emergent hierarchy to the videos, defaulting in some being selected and some left unseen. The outcome of these actions are then monitored to determine the score, with it going from low , not interacted with, to high reward, selected and viewed in full, to the train the agent further.

These three elements form how an agent is figured and what the agents are able to experience and do. Traditionally these bodies have been fragile, only being suited to particular tasks/problems, as well as not being very adaptive to new situations, but through adapting not only the architectures of their NNs but also the ways we figure the agents environments and how they are trained/monitored it has allowed for recent improvements within RL in the last few years. This is shown particularly in a paper/project by Yujin Tang and David Ha (2021) where RL models with their new sensory transformer layer can handle inputs randomly sorted in order with the models still working well. In

their online publishing of the paper here there is an interactive version using a cart-pole[2] example but modified with their new layer, and bellow is also a gif of a pong with the its graphics iteratively rearranged and even reduced to just 30% of the total output. Even though these examples are inspired by the human ability for sensory substitution[3], the power and adapt-ability of these agents is far outside of ours, can you imagine being able to play the game bellow, within such a fragmented and changing environment.



This ability of the NN from an anthropocentric position seems perplexing, but as these models already take in very abstracted and fractured sensing's of their world/environment, so why should this re-fracturing be any worse? When it comes to seeing these abilities in the context of this research (RS), it is exciting to think how mechanisms developed in the same trajectory might be able to imagine and navigate other dislocated terrains and matterings, such as archives, records, and their narratives, performatively traversing their complex and everchanging ontologies and taxonomies to develop and test out new boundaries, connections and forms of knowledge. This flexibility produces indeterminate assemblages showing hope for them as a collaborative tool to cross boundaries, fathom the unknown, and weave realities/narratives into dialogue.

The main question that the current circumstances prompts me to ask is how can we figure environments/architectures that promote these agents ability to navigate these spaces?

---

[2] A classic test for RL, balancing a pole vertically on a movable panel, learning how virtual gravity works.

[3] Sensory substitution refers to the brain's ability to use one sensory modality (e.g., touch) to supply environmental information normally gathered by another sense (e.g., vision). Numerous studies have demonstrated that humans can adapt to changes in sensory inputs, even when they are fed into the *wrong* channels [1, 2, 3, 4]. But difficult adaptations—such as learning to "see" by interpreting visual information emitted from a grid of electrodes placed on one's tongue [2], or learning to ride a "backwards" bicycle [3]—require months of training to attain mastery. Can we do better, and create artificial systems that can rapidly adapt to sensory substitutions, without the need to be retrained?

Whether it is how they are guided or how they are able to sense and do, this is a place where we may need to step out of our ways and into theirs (Bunz) to explore and feel potential figurations of RL and archives that make our relationships with intelligent agents' access-able to new kinds of knowledge.

2.3 Valuation (questioning the reward system and volumetric ecologies)

The very materiality of ML and NNs are valuing apparatus, the dialogues are through numeric valuations and representations of space and things, and the training dynamic is a stochastic valuation of actions and their results towards a mathematically constructed set of values. Within RL itself the reward or penal system are essentially landmarks of valuation for the model to learn from and develop around. Structures of valuation such as these are what De Silva examines in her essay "On Matter Beyond the Equation of Value", discussing how concepts of efficient causality played out through the historical evolution of European knowledge and how it has figured formals of modern understanding, legitimisation, and dialogue. De Silva in this essay and her greater work questions how we can rewrite the formulas we have been given to determine new and fruitful dialogues. At the end of the essay she presents an impossibility to these formulas by conjuring infinity from their deterministic and derivative clasps. Recomposing Black life not as an absence or negative of life but as a place of opportunity and possibility, presenting the infinite that a body can hold but a formula can't.

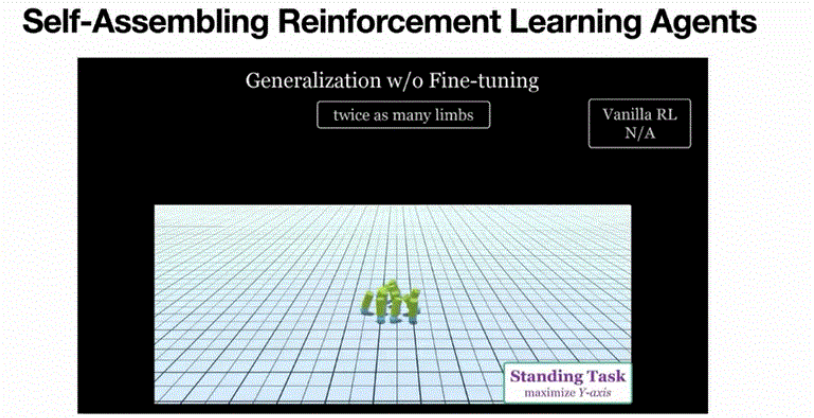
In a similar direction we can question how these valuation formulas are enacted or imposed within intelligent apparatuses, from the normalisation of data/input/output values to the regulation and reinforcement of behaviours through valuation. These are such complex and delicate apparatuses that it is hard to know directly where to intervene with these formulas, but we can practice to think how this might be done, and in which avenues we can form fractures within these formula to spring from them new infinities and abundance.

2.4 Continuity (questioning time and figurations)

RL agents can be used on a one-shot basis or emergently through a series of steps trying to achieve their goal, this cycle can also finish early if they fail at their task. Time itself isn't always explicit within an RL agent's environment but it can be sensed as an input if the environment needs it. For instance, chess does not inherently need time to understand what is going on, but potentially a recommender for a social feed might benefit it as an input. On top of this the sense of time can be abstract sentiments, such as the step number in the episode or scores (5==end). The one constant in training is that there is always an end to the episode, even if the task doesn't need one, the algorithm will inherently impose one and disrupt it to create episodes for the training to compare and improve upon. Once training is completed these artificial segmentations to a task can be removed/ignored and an agent is able to work continuously within its operations, only to be interrupted when replaced by another trained model or algorithm.

Multi-agent environments in many ways queer and crip time in several interesting ways as it doesn't have to have a classic linear cycle in any shape or form. Agents' have the possibility to collaborate with other models (NN) but it is often an instance of its own model reflecting back at it or a checkpointed version of itself from earlier in the training process. These checkpoints are described as taking on different personas and strategy types, expanding and testing the learning agents in diverse ways against a less erratic yet dynamic partner. In this figuration you see refracted selves from different times and places, learning from one another and building up a set of knowledges through playing together. At each step an environment has the potential for calling the same NN

from infinite number of instances, forming the phenomena and dynamics of the environment through its emergent relationships. This configures a RL's NN to be multi-bodied and situated, holding the capacity to transform and perform many roles and relationships within a dynamic emergent environment, able to transform its behaviour to adapt to changing terrain, differing perspectives and roles.



Deepak Pathak, Chris Lu, Trevor Darrell, Phillip Isola, Alexei Efros. Learning to control self-assembling morphologies: a study of generalization via modularity. NeurIPS, 2019.

The example above and below are explicit examples of this sort of queer and crip continuity, taking a cellular approach to problems provides a different solution to a task, breaking it down to have an agent perform each limb in the example above, and each pixels in the example bellow, with the NN taking multiple calls from agents in different situations and working together to resolve a robust and adaptive solution through each of these situated actions. In some settings these cellular formations have surpassed the abilities of their non-cellular counterparts, in particular showing great ability to deal with unexperienced phenomena by being adaptable and reconfigurable, whilst also working with smaller NN architectures that are faster to train.

Ettore Randazzo, Alexander Mordvintsev, Eyvind Niklassonand, Michael Levin, Sam Greydanus: Self-classifying MNIST digits. Distill, 2020.

Returning to the title of this section, "Continuity", it could be easy to say that these configurations have no sense or ability for continuity, with their fracturing cycles and perceptions, compounded through multi-positionalities. In response when you look a bit deeper, they have moved into a new sense of continuity, one away from the predispositions of linearity and into new forms of causality. Their continuity is one of a high dimensional entanglement, intra-acting and understanding situated cause and effect through many roles and positionalities. So what does this mean for us as at the handle end of this new spoon? Moving through these high dimensional and fluid sources of archives with RL as our new organizational partners, how can we work with them to sense anew and touch upon that we cannot yet figure.

Maybe talk about Margulis ?

2.5 Indeterminacy (questioning control/entropy, freer agents without deterministic task)

RL differs a lot from other ML figurations not only by its ability to learn from an infinite environment instead of curated data, but by its foundations as a seeded and randomized model, meaning that it is inherently indeterminate. This means that it is limited to exploitation for definitive tasks such as categorisation as its outputs may vary, creating multiple options for what it might be. For dynamic and emergent problems where there may be many answers or possible ways of doing things it is a perfect adaptive and creative solution. When testing RL agents it has to be done by comparing multiples of runs alongside one another, seeing how their choices/paths converge in a similar successful paths, or if they diverge into diversely different modes. When in real world situations these actions can be averaged out over many calls of the same step (input data) to create a less sporadic averaged action (output), but it is still resting on and coming from a path of many, a rope woven tight.

Reflecting back to the archive and recommender systems that have brought us here, these properties and behaviours make RL the perfect collaborator into our gapped and tangential formations of knowledge, creating a way to fabulate a plurality of paths to bring these disparate and isolated (within the western tradition) phenomena together through a web of multi-poss-ability. Moving us away from a monolithic determinate and authoritarian narrative, and into a multiple of narratives rooted from potentially dynamic and situated agents, bringing many stories and worlds to explore and find ourselves, instead of the one we compress ourselves into.

2.6 Languages (Questioning collaboration and communication)

This essay has in many ways talked a fair amount about NN's languages, but potentially not so intrinsically. The RL's input, output and reward being their basic modes of dialogue/communication, forming from these exchanges with their world more complex patterns or ph(r)ases are formed/recognised, and in turn more defined wishes and actions can be cast into their emergent environment. Other sorts of ML could also be navigated similarly, with for example an image of a dog being shown to a compute vision model with it then communicating the defined number (not word) for dog, but for me the ability of RL to understand and navigate emergent problems means these notions of language are more adaptive and evolving in comparison to static supervised models.

The importance of NN's languages when used in this way does not stop at the black box dialogues it enables, but extends into the granular body of the NN. These models work emergently from nodes of statistical graph functions, creating layers of curves that filter and transform the input into the output, creating in effect a neural cascade through tuned statistical relationships. NN's body is one of relationships and movements, nodes in dialogue, it could easily be said that each of these interconnections within a cascade has its own situated language, sensing in and communicating out, forming internations of happenings and composing symbols and signifiers. These granular syllables combine and fold through each layer of the cascade, fabulating worlds from the emerging syllable-s-word-s-sentence-s-scene-s-image-inary-s, forming a knot to rest on or a superposition to come from. This way of forming an emergent language is an adaptive fairly open ended one, with the ability to create a highly dimensional way of navigating a problem, this dimensionality reflects our languages

within words and numbers, but within these classic ways of wor(l)ding we seem to get stuck in low dimensional dialogue, squeezing complex problems through binary set narratives, figurations and infrastructures. Returning to the CV model sensing a dog, its only output is through singular indexes or a limited set of categorical statistics that can only navigate a select few defined outcomes. NNs form in an oppositional direction, from the granule up, allowing agents to build up and sediment fluid words that they use to describe their world, environment and actions, with thousands of dimensions to a single form.

This figuration of their language has been something that has perplexed many statisticians, data scientists and mathematicians, as in many ways we cannot translate NN's languages or transformations into our own. We can in many ways navigate it from a higher level (black box) where we can talk about behaviour and these sorts of dialogues, but these are again the levels that we are comfortable navigating but aren't always as clear and certain as many would present them to be. Within the blossoming field of explainable AI (XAI) there have been studies that show that what 'experts' think models are doing and what they actually do can be vastly different, making them still elusive to all levels of research. This is exacerbated by the granular level of NN's statistics, and this is where we become distanced, and incommunicable. You may ask why? These languages are statistic models that we created that use fairly simple curve graphs to explore problems, surely their means are the same as ours, surely this is our language that they learn to use and which we can navigate? But no!

This is one of the most exiting parts of NN for me as even though we have formed their body and commence their mattering's, their nature is still something we cannot navigate or explain. The emergent and layered languages that they have formed and composed are ones that actually defy ours mathematic and statistical traditions, they can easily break the axis of the graph (something that is impossible in classic human graph relations) and still form complex and accurate descriptions of phenomena. Much like Barad and Bohr's interpretation of the twin split experiment, or De Silva's infinity, within this "impossibility" are new (pos)abilities, what if we are too stuck within our modes of language to see or understand new ones[4]? This question is always a pertinent one within feminist critiques, but this not so recent step brings it to the surface once again.

RL has taken this notion of language further within multi-agent environments by their ability to form languages between agents that can transfer messages between each other to help them collaborate more cohesively. These evolved signalling dialogues are simply extra inputs and outputs (actions) that each agent can learn to use and understand over time to be able to form collective movements. Many of the earlier mentioned RL examples in fact use this ability to their strength, the categorisation is the most simply understood as it is cellular and grided, with each agent/pixel talking to its neighbours, not only through its selection, but also through unseen parameters. Beyond this example it is also their ability to negotiate and translate these internal and dimensionally dense milieu of feelings/words/signals, that can be openly translated/communicated between agents and collectives.

- Maybe talk about the neural re-presentation work I did?

Languages and their logics are key to these models functionality, impressively forming from relations instead of mapping to them, these notions are inspirational ones not only for figuring models but

[4] Ref on how animals can understand us but we cant understand them…

also for practice, how can we build up instead of project down, and navigate change, instead of undermining it.

2.7 gradient descent (questioning navigation)

gradient descent is the mechanism of most NN's training, moving a randomised net towards a refined solution through iterative landmarking and tweaking of weights and biases. On a metaphorical level it is a performance of scientific practice played out through a cyclical algorithm, testing, sensing, and using formula to project outcomes and ontologies to then make actions to improve its theorem. The fact is that science isn't as "pure" as its imaginary is, and social biases and agendas tend to obscure the smoothness and effectiveness of these cycles, and the same is true for theses models. Not only are the volumetrics/parameters they use constructed around biased data, or in RL's situation Thorndike's penal system or the bias environments constructed through it, but also the guiding mechanisms of gradient decent and training, loss functions; as embedded in them are segregationist ideals and monolithic embodiments(Chun). One normalizing and centralized function is used to define the difference between the current prediction and the "truth" to determine which way the nets weights and biases must be tweaked. These are often static function, but in recent years there have been dynamic implementations that allow for loss functions to adapt to the model that is being trained. (maybe add more about how it pluralises prediction)

The loss from a supervised/unsupervised models training will be quite linear, with them being applied repeatedly to the same set of data and relationship, reducing its loss through steps, aiming for 0. RL queers and crips this linear perception of loss through the unknown, using experience, change and revelation to diverge a path to it's own "truth". Their paths feel quite relatable, eagerly thinking you may understand a place or process through a few early experiences with your confidence gained and your loss low, but when you find something new or experience a moment that makes you question it again, setting a new standard or introducing an unknown problem, you inevitably loose confidence in your ability and your loss rises. The key to this process is the repetition cycles of building up If we are to relate to this queered loss, for me it is to know that confidence is fluid and a spectrum, it tells you when you when to

Loss function into the unknown

3 Conclusion

Throughout section two of this paper I have explored a few gateways into, and axis of RS/RL, trying to navigate a new understanding of how this apparatus ids understood and navigated. It is not my intention to tie these elements up into a definitive new modelling, definition or direction, but instead leave them open to be reinterpreted and renegotiated. This action in itself