Empowerment and Relevant Goal Information as Alternatives to Graph-Theoretic Centrality for Navigational Decision Making

Marcus Clements
Adaptive Systems Group
University of Hertfordshire
mc@mxcog.com

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of MSc by Research, September 2019

Abstract

City planners and architects employ graph-theoretic measures to analyse models of the built environment and predict human navigational behaviour. A recent breakthrough in the neuroscience of spatial cognition has shown that activation in the human hippocampus tracks the change in centrality for subjects navigating a virtual Soho in a fMRI scanner. Based on a well understood information-theoretic framework for modelling intelligent behaviour under cognitive constraints, the existing measures *empowerment* and *relevant goal information*, and novel quantity *relevant goal information uptake* were applied to a graph of the Soho street network navigated in the experiment. Empowerment, relevant goal information and relevant goal information uptake are shown to correlate with graph centrality for the primal graph, and to a lesser extent with centrality for the dual graph as used in the Soho experiment. These results, consistent with the hypothesis, provide preliminary evidence that human navigation employs an empowerment maximisation strategy, and to the author's knowledge, linking empowerment and relevant goal information with empirical neuroscience in a collaborative study for the first time.

Acknowledgements

I would like to thank:
Prof. Hugo Spiers (University College London), Dr. Amir-Homayoun Javadi (University of Kent) and Dr Joao Pinelo Silva (University of Bahrain) for their help and data for the Soho navigation experiment.
Prof. Daniel Polani (University of Hertfordshire) – I can't imagine a more inspiring and engaged supervisor. We covered so much ground, usually goal-relevant, always informative, and certainly empowering.
Dr. Volker Steuber (University of Hertfordshire), my second supervisor.
My colleagues in the Adaptive Systems Group, and the Sepia Lab especially for welcoming me into a friendly, diverse and intellectually stimulating community.
Dr. Nicola Catenacci Volpi (University of Hertfordshire) for his patience and help with learning the mathematics.
The people who made me and raised me: Karin Clements, Robert Clements and Deborah Clements, who have had to wait for a long time for me to get letters after my name.
My wonderful partner Vahsti Hale, without her love and support this would have been so much harder.

Contents

1 Introduction

When asking directions to a destination in a city it is common to say "What is the easiest route?". The adjective "easiest" encapsulates the optimum blend of a set of competing priorities. Is it the route with the shortest distance, the route with fewest junctions, or the route that passes through the best connected streets? When navigating, humans choose a route that combines these and other factors, balancing the trade-off of physical effort, cognitive effort and travel time.

A substantial body of work at the intersection of artificial intelligence and cognitive science, developed by Daniel Polani's Sepia Lab in the Adaptive Systems Group at the University of Hertfordshire, has provided information-theoretic formalisms for the quantification of cognitive effort and constraints. These formalisms have been applied to analyse models of intelligent behaviour including navigation (Polani et al., 2001; Klyubin et al., 2005*b*,*a*; Polani, 2009; Van Dijk and Polani, 2013; Salge et al., 2014*b*).

City planners have long depended on the graph-theoretic analysis of street networks to predict pedestrian and traffic flow (Hillier and Hanson, 1984). Influential work on the neuroscience of spatial cognition by Hugo Spiers' Lab, in the Spatial Cognition Group at University College London, has shown that activation in the right posterior hippocampus, a brain region associated with navigation, scales in proportion to changes in graph centrality (Javadi et al., 2017).

The aim of this work is to develop a simple artificial intelligence agent model of human navigation, and to compare the information-theoretic predictions for navigational behaviour from the model with a graph-theoretic analysis of the topology of the street network used in the Soho navigation experiment reported in (Howard et al., 2014) and (Javadi et al., 2017). I argue that the information-theoretic treatment is more general than the graph-theoretic because it is grounded by the concept of organisms as information-processing entities. Where graph theory deals specifically with the abstract representation of pairwise relationships, information theory has significantly more general applications and provides tools for analysing the flow of information that naturally extend beyond the simple discrete model described in the present work, to continuous domains and stochastic environments more like the real world. This study takes the first step towards demonstrating that aspects of human navigational behaviour can be predicted by the information-theoretic measures *empowerment* and *relevant goal information* (detailed in Section 4). The results (Section 5) show that, consistent with my hypothesis (Section 1.7) that when applied to a model of an agent navigating in the Soho street network, *empowerment* and *relevant goal information* correlate with the graph-theoretic measures *degree centrality* and *closeness centrality* used in the Spiers Lab study. Furthermore the results show that change in empowerment correlates with the change in degree centrality, which was found to correlate with activation in the right posterior hippocampus in Javadi et al. (2017). Having provided indirect evidence of the ability of these measures to predict brain activation through the correlation with graph centrality, the present study provides a foundation for a direct comparison with fMRI brain activation data in future work.

In the remainder of this section there is a brief review of the literature from psychology, probability, information theory, neuroscience, spatial cognition and artificial intelligence that underpins the present work, before outlining the hypothesis and contribution to knowledge in Sections 1.7 and 1.8.

1.1 Organisms and information

The principle of natural selection, via competition for finite resources, ensures that living things have evolved effective strategies to maximise the chance of propagating their genetic code to future generations (Darwin, 1859). An effective strategy for life must optimise the acquisition and utilisation of energy. In animals, most of the energy consumption in the body is due to maintaining homeostatic functions and movement, but a significant proportion, is consumed by the sensory systems, nervous system and brain. In resting humans up to 20% of total energy used is consumed by the brain (Kandel et al., 1991). Humans have evolved such expensive brains for good reason – the subtlety and complexity of the strategies that the human brain may acquire and execute, have allowed us to adapt successfully to live almost everywhere on our planetary landmass, with ever increasing lifespans and population size.

In 1948 communications scientist Claude Shannon published his seminal work, developing the mathematical tools of information theory, to quantify information over noisy communication channels (Shannon, 1948). Central to Shannon's work is the concept of *entropy*, a measure of disorder that quantifies the information that may be transmitted over a channel. Informational entropy has an analogous concept in thermodynamics (after which it was named), and the relationship between informational and thermodynamic entropy, implies a connection between Gibbs free energy and the amount of information processable by a system (Jaynes, 1957*a*,*b*; Polani, 2009).

The idea that perception depends on a probabilistic mechanism can be traced back to Hermann von Helmholtz, who described visual perception as a process of "unconscious inferences" (Von Helmholtz, 1867). The concept that information is acquired through the senses and processed by the nervous system and brain, is rooted in work by McCulloch and Pitts (1943), extended to a formal *computational theory of mind* by Hillary Putnam and Jerry Fodor in the 1960's, 70's and 80's (Rescorla, 2015). A central concept enabling the analysis of cognition with information theory is the *perception-action loop* which describes the *cybernetic cycle* of information from the environment through an agent's sensors, and back via the agent's actuators. This concept, based on early work by William T. Powers (1973), gained popularity in psychology, psychiatry and neuroscience literature during the 1980's (Fuster, 1990). The quantitative analysis of perception and action through information theory was suggested by Ashby (1956), developed in work on control theory by Touchette and Lloyd (2000), and linked to utility theory with the concept of *relevant information* (Polani et al., 2001). The relevant information formalism, detailed in Section 2.4.1, from which *relevant goal information* that is used in the present study is derived, enables us to quantitatively answer the question "What is the minimum information required to take action?", and to model optimal decision making in terms of information processing cost via Shannon's tools.

Considering the perception-action loop to be a communication channel, enables the quantification of the information flow from the environment into the organism via sensors and back out again via actuators. The *channel capacity* (Shannon, 1948) of this actuation channel is known as *empowerment* (Klyubin et al., 2005*b*,*a*), because it measures the perceived potential for an agent to influence the world, either by changing its own state, or by changing the environment (Salge et al., 2014*a*). Empowerment has been computed in discrete and continuous models (Jung et al., 2011; Salge et al., 2013). Empowerment can capture properties of the environment, and can be employed as a task-less intrinsic motivation for agents that maximises opportunity. Salge et al. (2014*b*) state the *behavioural empowerment hypothesis* as "The adaptation brought about by natural evolution produced organisms that in absence of specific goals behave as if they were maximising their empowerment". Skiers and snowboarders understand the concept of maximising their empowerment

due to gravity, staying high until the goal is clear because a premature descent results in a gruelling trudge. The empowerment formalism is detailed in Section 4.1.

The quantity of information that an organism can acquire through its sensory machinery is limited by physical and biological constraints, and represents a small fraction of the information available in the environment. A central theme of the present study is the *information parsimony principle* (Polani et al., 2007*a*) which holds that organisms must have evolved sensors and neural architecture to capture only the most *relevant information* (Polani et al., 2001), and to capture an amount of information, where the costs of acquisition and processing are optimised with respect to other energy demands. In other words, an organism must "trade-off" informational costs, against other energy costs (e.g. from movement), to best exploit the opportunities provided by the world (Polani, 2009).

This trade-off means that human decision-making is not always "rational" – cutting cognitive corners inevitably leads to inaccuracies and biases. Cognitive constraints as the source of bias in economic decision-making were investigated by Herbert Simon from the 1950's (Simon, 1957, 1978), and the study of cognitive biases has continued to expand in psychology and behavioural economics, including influential work by Amon Tversky and Daniel Kahneman (Kahneman and Tversky, 1979; Kahneman et al., 1982; Kahneman, 2011). Lending weight to the information parsimony hypothesis, aversion to cognitive demand was demonstrated experimentally by Kool et al. (2010). The information-theoretic analysis of the perception-action loop models influence of cognitive constraints on decision-making performance in work by Van Dijk and Polani in the early 2010's that provides the theoretic basis for the present study(Van Dijk et al., 2010; Van Dijk and Polani, 2011, 2013).

Much of the present work depends on two cornerstones of artificial intelligence theory, the *Markov decision process* (MDP) and *Reinforcement Learning*. Richard Bellman developed *dynamic programming* for MDPs (Bellman, 1953), and Ron Howard first described the policy evaluation method described in Section 3.3 (Howard, 1960). For a comprehensive review of dynamic programming see Bertsekas (2005). Reinforcement learning for artificial intelligence was first developed by Andrew Sutton and William Barto in the early 1980s (Sutton and Barto, 1998). A sizeable body of work relates the formalism of reinforcement learning to neural activity to determine if, and where, values, rewards, policies and transition-models are stored and accessed in the brain of humans and animals (Botvinick et al., 1995; Lee et al., 2012). Reinforcement learning models of decision making and learning have, in turn, elucidated empirical work identifying brain regions and suggesting neural correlates for reward-based learning and goal-based decision making (Solway and Botvinick, 2012).

Agents in a *model-based* reinforcement learning experiment may begin with a model of the dynamics of the world, known as the *state-action transition function* (see Section 3.3), and they must learn a *policy*, typically by maximising some reward function, which enables them to select actions according to state. Model-based decision making has been studied by psychologists for a long time, including the concept of the cognitive map (Tolman, 1948), which may allow mammals to switch policies easily as goals and rewards change. More recently Penny et al. (2013) propose a dynamic Bayesian model of model-based planning and spatial cognition, mapped onto the function of the entorhinal cortex and hippocampus.

In contrast, for *model-free* experiments, the dynamics of the world is not already known to the agent, but it may learn a policy nonetheless, for example via *Q-learning* (Russell and Norvig, 2009). Differences in brain activation under these two regimes have been demonstrated in fMRI studies of humans navigating (Hartley et al., 2003). In the Spiers Lab experiment described in

Section 1.4 (Howard et al., 2014; Javadi et al., 2017), subjects were extensively trained on the area of Soho in which they would be virtually navigating – by the start of the experiment they had already acquired an accurate cognitive map. The participants were not expected to learn about the environment during the experiment, justifying the use in this work of a model-based approach, where the transition function is already known (Section 3.3).

Relevant information (Polani et al., 2001), unlike earlier attempts to assess the value of information (Howard, 1966), provides a method of evaluating the pertinence of sensor information by measuring the minimum sensor information required to achieve a certain level of performance (utility). The extension of the relevant information formalism to goal-directed activity is termed *relevant goal information*, which measures the minimum amount of information *about the goal*, under a similar utility constraint (Van Dijk et al., 2010). An agent acting in such a way as to minimise the goal relevant information needed to take action on average, will choose a routes that are most likely to be shared across goals, with beneficial implications for spatial cognition in terms of memory and processing. An agent following an empowerment maximising strategy will also tend to navigate through the same states for multiple journeys, however empowerment is a local quantity with a limited horizon, whereas relevant goal information is a global quantity and it the two measures uncover different properties of the environment.

The information-theoretic measures of graph topology utilised in the present study are *empowerment*, *relevant goal information* (RGI), and the novel extension of RGI (based on a similar quantity formulated by Van Dijk (2013)) to include knowledge about the history of the agent dubbed *relevant goal information uptake* (RGIU). These measures are detailed in Section 4 with data for a grid world model of interior space, and data for the Soho street network from Javadi et al. (2017) used as illustrative examples.

1.2 Embodied cognition and the Bayesian brain

Embodied cognition, or *embodiment*, is the theory that cognitive processes, perception, and intelligence are tightly bound to the physical organs that enable their function, and these sensory organs are in turn morphologically coupled to the environment. The mind evolved with the body, and intelligence is the outcome of the development of embodied cognitive processes, optimising the processing of information by exploiting the structure of the environment. By considering life as information-processing it becomes difficult to define where the processing starts and ends in the perception-action loop of world state information, perceived state information and actions and decisions. Is the brain the information-processing unit? Should we include the nervous system as a whole? Or extend the definition to the sensors? The shape of the human ear filters sound to improve intelligibility, pre-processing auditory information before it reaches the timpanum. The embodiment of an organism induces a structure in the flow of information into the brain, conferring an advantage by processing that information to be more salient to the organisms needs, reducing the demand on the neural part of the chain. The causal Bayesian model of the perception-action loop, as applied in the present study, formalises the embodied dynamics of information as a family of probability distributions, permitting the quantitative analysis of the information flow between an agent and the environment (Polani et al., 2007*b*).

Bayesian models have been successfully applied to brain activation, neural activity, and sensorimotor processes, and are proposed as the "underlying computational infrastructure of the brain" (Penny, 2012). Besides the formalisms derived from the interpretation of the perception-action loop as a causal Bayesian network (Section 2.4), a number of other approaches have been taken

to modelling fundamental properties of intelligent behaviour with Bayesian models, revitalising the Helmholtzian view (Dayan et al., 1995). The formalism of predictive information (Bialek and Tishby, 1999) has been used in artificial agents as an intrinsic utility function which exhibit various "intelligent" behaviours (Ay et al., 2008). The *free energy principle* (Friston et al., 2006), describes how organisms try to reduce "surprise", by minimising the difference between their model and the world as perceived, or by acting to change the world to more closely match the predicted model (Friston, 2009, 2010). Work by Friston and others on free energy and active inference, has formed the basis of a large body of literature in computational neuroscience, experimental psychology, and philosophy, including the predictive processing theory of mind (Clark, 2013). Recent work suggests that theses strands of research have much in common - empowerment, predictive information and free energy are different ways of formulating the Bayesian minimisation of the difference between the posterior and prior beliefs about the world, and can be reformulated to emphasise their similarity (Biehl et al., 2015, 2018). Bayesian models of embodied cognition, via the reduction of sensory uncertainty by predictive processing, have been extended to interoception, emotion and the sense of self (Seth, 2013). Other researchers have investigated information-theoretic explanations for neural activity, in experiments by Strange et al. (2005) the anterior hippocampus was found to be sensitive to the entropy of visual stimulus.

The concept of *intrinsic motivation* describes how an agent can be driven to behave in the absence of external goals. Computational approaches to intrinsic motivation have given artificial agents the ability to exhibit explorative curiosity (Schmidhuber, 2010), an important component in intelligent behaviour. Agents motivated to maximise their empowerment, an information-theoretic approach to intrinsic motivation employed in the present study, exhibit intelligent-like behaviour in a number of settings as individuals (Salge et al., 2014*b*), and collectively (Capdepuy et al., 2007, 2012; Clements and Polani, 2017). Other related variational information approaches to intrinsic motivation include *infotaxis* (Vergassola et al., 2007) and *infotropism* (Thornton, 2014). Free energy maximisation also provides a mechanism for the intrinsic motivation, with the benefit of biological plausibility when used as an objective function for active inference. Biehl et al. (2015) present a unified mathematical treatment of active inference with alternative objective functions including empowerment. Research into intrinsic motivation may also play a role in the development of *artificial general intelligence*, perhaps in helping solve the famous *value alignment* problem (Salge and Polani, 2017).

1.3 Spatial cognition

The first moving organisms may have appeared as long ago as 2.1 billion years ago (Albani et al., 2019). For almost all animals, including humans, the ability to successfully navigate the environment to find food, or prospective mates, is of paramount importance. In the mammalian brain, the hippocampus plays a central role in navigation, and there is a growing body of evidence that the same neural structures that have evolved to support navigation also support abstract spatial representations (Epstein et al., 2017; Garvert et al., 2017). The role of brain regions, primarily used for spatial cognition, in other aspects of decision making is hinted at by our use of spatial language when describing problem solving in other domains e.g. "Finding a solution", "Project milestones", "Development road map". The concept of the *cognitive map* was first suggested by Tolman (1948), and located in the hippocampus with the discovery of place cells (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978).

In a study by the Spiers Lab at UCL, human subjects were placed in a fMRI scanner and asked

to navigate to a familiar location in a video simulation of a walk through the streets of Soho. The hippocampus was found to track spatial parameters including *egocentric goal direction*, as well as *path distance*, and *euclidean distance* to the goal (Howard et al., 2014). In a further study using the same experimental data, activation in the right anterior and posteror hippocampal regions were found to correlate with changes in *degree centrality* and *closeness centrality* respectively (Javadi et al., 2017). The centrality data from (Javadi et al., 2017) forms the basis of the collaboration reported in the present study.

The centrality measures used in Javadi et al. (2017) were computed using a suite of methods known as *space syntax*, pioneered by Bill Hillier and Julienne Hanson in the 1980's (Hillier and Hanson, 1984), that are commonly used to predict pedestrian and vehicular flow, in planning the built environment (Al_Sayed et al., 2014). The three measures used – *connectivity*, *choice* and *integration*, are closely related to *degree centrality*, *closeness centrality* and *betweenness centrality* respectively. Betweenness centrality was first formalised by Linton Freeman in the study of social networks (Freeman, 1977), Brandes and Erlebach provided a comprehensive review of centrality measures including degree centrality and closeness centrality (Brandes and Erlebach, 1998).

Does the hippocampus directly encode a measure of centrality, or can an information-theoretic treatment of cognitive constraints shed more light on the correlation between centrality and brain activation? A cognitive explanation of the correlation of graph-theoretical space syntax quantities with human urban movement pattern has been suspected for some time (Hiller and Iida, 2005). Masucci et al. (2009) describe an "interaction between the metrical and informational space" and how "a principle of least effort" explains human urban navigation, hinting at the information parsimony principle (Polani et al., 2007*a*), but they do not apply Shannon's tools to the perception-action loop in their analysis.

1.4 Soho navigation study

The Spiers Lab navigation study used ten different video routes through Soho. Each route has an origin somewhere along a street, and ends on another street after a series of junctions. A cohort of 25 participants were well trained in navigating the local street network, and their knowledge of the map was tested and confirmed before beginning the experiment. Results have been published in several papers including Howard et al. (2014) and Javadi et al. (2017).

During the experiment participants, prone in a fMRI scanner, were shown videos from a first-person perspective of a journey by foot through Soho. Each route begins with the participant being shown the name and photo of the goal for 8 seconds. The video then proceeds to follow the shortest path to the goal. At certain "decision" points in the journey, the video stopped and they responded to navigational questions, such as "Goal L/R?", by pressing "Left" or "Right" buttons on a hand held controller. Each participant was shown a video for one *navigation* route, where they were asked to think about navigating the optimal route, and had to answer the navigation questions, and one *control* route where they were asked not to think about navigation or way-finding, and no questions were asked. Navigation and control routes were allocated to participants at random. At certain junctions the video does not follow the optimal route, but instead a detour from the optimal path is taken, so the participants were forced to re-plan before responding at the next decision point.

The fMRI data was continually recorded, and the spatial measures in the analysis were compared with brain activity at key events including *junction entry*, *street entry* (junction exit) – when a turn through a junction had completed and the virtual journey had entered a new street, *decision*

points, and *travel points* at randomly jittered intervals during traversal of streets.

The main results reported in Javadi et al. (2017) were:

1. The change in degree centrality ΔC_D is tracked by the right posterior hippocampus during street entry events.

2. The change in closeness centrality ΔC_C is tracked by the right anterior hippocampus during street entry events.

3. Activity in the lateral prefrontal cortex scales with demands of a breadth-first search at forced detours.

1.5 Space Syntax and the dual graph

As described, in Section 3.2.1, street networks can readily be represented by a *primal* graph where the vertices are junctions and the streets are edges. Space Syntax however, deploys the *dual graph*[1] where the vertices are streets and the edges are junctions. This representation is described as the "information-space" of the street network (Masucci et al., 2014), and there is a considerable body of work considering the application of the dual graph to street network analysis and urban planning (Porta et al., 2006; Masucci et al., 2009, 2014). The dual representation lends itself well to analysis of street names because several segments can be connected together as a single vertex. The angles between the streets may be represented by edge weights in the dual representation allowing vertices (street segments) to be merged according to *axial lines*. Both these ways of compressing the state space in the dual representation by merging street segments are important concepts in city planning (Hiller and Iida, 2005). The measures described in Javadi et al. (2017) are street segment values – in other words, centrality of the vertices in the dual graph. It is beyond the scope of this work to consider the dual graph. Applying the information-theoretic treatment to the dual graph representation of streets and junctions requires significant further theoretical study and evolution of the formalisms, and as such is left to future work.

Despite the success of Space Syntax, the dual graph is not universally accepted as more suitable for urban spatial analysis than the primal graph – the primal graph is more suitable for representing the street network when used alongside other spatial and geographical features including length, width, circuitry and position of buildings (Boeing, 2017). Taking these advantages into account, it seems reasonable to propose that the primal representation may share enough with the structure of spatial information in the human brain, to complement the dual graph results in Javadi et al. (2017), in terms of brain activity correlation during navigation.

1.6 Aims and method

One of the most exciting outcomes of the study of artificial intelligence is how the models developed can advance our understanding of animal and human intelligence. Artificial agents employing policies utilising empowerment and relevant goal information have been shown to exhibit intelligence-like behaviour in a range of modelled environments. The future goal towards which this work takes the first steps, is to connect artificial intelligence with empirical neuroscience by testing whether animals and humans employ an empowerment-like, or relevant goal information minimisation strategy during navigation.

[1] This is a special form of dual graph known as the *edge-vertex dual graph*, or *line graph*, referred to in this text as *dual graph* for consistency with the city planning and architecture literature.

Graph-theoretic tools such as Space Syntax have been shown to predict human navigational behaviour in cities and architectural spaces. If the information-theoretic quantities measured in these environments modelled with a graph are correlated with the graph-theoretic quantities, then there is a sound basis to investigate a direct correlation between the information-theoretic measures and brain activation in the Spiers lab Soho navigation study.

I deploy a simple discrete probabilistic agent-based model, widely used in the study of artificial intelligence, of an agent navigating interior space and apply it to street networks via a shared representation of the agent's environment and dynamics as a graph. The present study consists of a series of parameterised experiments computing the information-theoretic quantities and the graph-theoretic quantities for each scenario modelled. The results show in tabular form the correlation between the information-theoretic measures and the graph-theoretic measures.

1.7 Hypothesis

I hypothesise that the information-theoretic quantities *empowerment* and *relevant goal information* are correlated with the graph-theoretic measures *degree centrality*, *closeness centrality* and *betweenness centrality* when applied to graph representations of a model of interior space and the Soho street network.

1.8 Contribution to knowledge

The information-theoretic measures *empowerment*, *relevant goal information*, and *relevant goal information uptake*, based on a simple artificial intelligence agent model, are correlated with the graph-theoretic measures *degree centrality*, *closeness centrality* and *betweenness centrality* in a typical grid world used in the study of artificial intelligence, and in a graph representation of the Soho street network. Correlation with these quantities provides indirect evidence of the ability of the information-theoretic measures to predict navigational behaviour via prior work in city planning, and implies a link to hippocampal brain activation during navigation via an empirical study Javadi et al. (2017).

This work connects the information-theoretic study of artificial intelligence and agent-based modelling with empirical neuroscience. To my knowledge this thesis presents the first evidence for the connection of the information-theoretic quantities *empowerment* and *relevant goal information*, (via correlation with urban planning graph centrality measures), with human hippocampal brain activation during navigation.

The correlation with graph centrality for the street networks in Section 5 suggest that empowerment and relevant goal information could provide new insights into human navigational behaviour in urban environments for city planners, and the application of the measures to a model of internal space suggest applications for architecture and building management. The model presented here based on a graph representation unifies these applications into a single framework for analysis, providing a foundation for future work in both domains.

Relevant Goal Information Uptake (RGIU), described in Section 4.3 is a novel quantity developed here based on Sander Van Dijk's concept I^{new}(Van Dijk and Polani, 2013) represents the uptake of new information about the goal required by the agent in a given state. RGIU is for a limited number of steps of history, whereas I^{new} considers the full state history.

1.9 Structure of chapters

Section 2 provides an overview of the concepts from *information theory*, *graph centrality*, *Markov decision processes*, *reinforcement learning* and *relevant information*. The discrete probabilistic agent-based model, familiar from the artificial intelligence literature and deployed in the experiments is detailed in Section 3. Section 4 develops the concepts outlined in Section 2 in the context of the experimental model, and provides the derivation and interpretation of the information-theoretic measures *empowerment*, *relevant goal information* and *relevant goal information uptake*. The results consisting of a comparison of the information-theoretic measures with graph centrality are presented in Section 5. Lastly, in Section 6 I briefly discuss the findings and suggest directions for further research.

2 Technical Background

2.1 Information Theory

This section largely follows Shannon's original work (Shannon, 1948), and the comprehensive review of information theory by Cover and Thomas (Cover and Thomas, 2006).

2.1.1 Notation

The probability distribution of a random variable X with alphabet $\mathcal{X}$ from which an event x is drawn is $\Pr(X = x), \forall x \in \mathcal{X}$, replaced in this text by $p(x)$ for readability, when the random variable is obvious. By similar abuse of notation, where the probability distribution of X is conditioned on a specific value y of a second random variable Y, $\Pr(X = x|Y = y), \forall x \in \mathcal{X}$ is written $p(x|y)$. A sum or product will omit the set from which values are drawn, the full alphabet is implied e.g. $\sum_{x \in \mathcal{X}}$ is replaced by $\sum_x$. The cardinality of $\mathcal{X}$ is written $|\mathcal{X}|$. The expectation of a probability distribution $p(x)$ is written $\mathbb{E}[p(x)]$.

2.1.2 Entropy

The entropy of a random variable $H(X)$ quantifies the uncertainty about the value of X, and is the minimum descriptive complexity of X. The information contained in a random variable, can be considered to be the reduction in uncertainty that comes with knowing the value of the variable. Since I use the binary logarithm throughout, entropy and information are measured in *bits*.

$$H(X) = -\sum_x p(x) \log p(x) \tag{1}$$

Introducing a second random variable Y with alphabet $\mathcal{Y}$, the uncertainty about the value of Y given X is the *conditional entropy*.

$$H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) \tag{2}$$

If X and Y are independent, knowing the value of Y does not reduce the uncertainty about X, $H(X) = H(X|Y)$. Conversely if knowing the value of Y allows perfect prediction of X then $H(X) = 0$.

2.1.3 Mutual Information

The *mutual information* $I(X;Y)$ can be thought of as reduction of uncertainty in one variable provided by knowledge of the other, and quantifies the communication rate of a channel with transmitted signal X and received signal Y.

$$I(X;Y) = H(Y) - H(Y|X) \tag{3}$$

Properties of mutual information:

- symmetric $I(X;Y) = I(Y;X) = H(X) - H(X|Y)$
- non-negative $I(X;Y) \geq 0$

- zero if and only if X and Y are independent $I(X;Y) = 0 \iff H(Y|X) = H(Y), H(X|Y) = H(X)$
- bounded by the entropy of each variable $I(X;Y) \leq H(X), I(X;Y) \leq H(Y)$.

With the introduction of a third variable Z with alphabet $\mathcal{Z}$ we can consider the *conditional mutual information*, $I(X;Y|Z)$ which can be thought of as the reduction in uncertainty that knowing the value of Y provides about X, given the value of Z.

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) \tag{4}$$

2.1.4 Communication over a channel

A memoryless communication channel, without feedback, is modelled by the random variable X as the transmitted signal, with alphabet $\mathcal{X}$ distributed according to $p(x)$, and the received signal Y with alphabet $\mathcal{Y}$ distributed according to $p(y|x)$.

$$X \xrightarrow{p(y|x)} Y$$

With a perfect, noiseless channel, there is no uncertainty about the received signal $H(Y|X) = 0$, hence the mutual information is the full information of the transmitted signal $I(X;Y) = H(X)$. At the other end of the scale, a channel where the received signal bears no relation to the transmitted signal there is no mutual information $I(X;Y) = 0$. The trade-off between distortion and bandwidth is formulated both ways by Shannon as the *variational information* problems, *rate distortion* and *channel capacity*.

2.1.5 Rate Distortion

The rate-distortion formulation permits us to calculate how little information can be transmitted in this channel without distorting the original signal too much by quantifying the cost of mistakes as the *distortion* D_p caused by $p(y|x)$, and placing an upper bound on the distortion D_{max} – the maximum allowed distortion (or cost), over all possible valid probability distributions of $p(y|x)$:

$$\min_{p(y|x)} I(X;Y), \text{ subject to : } D_p \leq D_{max} \tag{5}$$

The distortion, or the cost of transmitting a symbol x and receiving y is commonly represented by a fixed distortion matrix $d(x,y)$, that does not depend on the dynamics of the channel $p(y|x)$. In this case, the total cost D_p is the expected value of the distortion matrix over the channel for all x and y.

$$D_p \coloneqq \mathbb{E}[d(X,Y)] = \sum_{x,y} p(x)\, p(y|x)\, d(x,y) \tag{6}$$

The convex constrained optimisation problem formulated in Equation 5 can be solved by the use of Lagrange multipliers. The first step is to convert the problem into an unconstrained problem by adding the mutual information term, which is to be minimised, to the distortion term, which is constrained, multiplied by a Lagrange multiplier β. There is now a new problem with the Lagrange function $\Lambda(p(y|x))$:

$$\min_{p(y|x)} \Lambda(p(y|x), \beta) = \min_{p(y|x)} I(X;Y) + \beta \mathbb{E}[d(X,Y)] \tag{7}$$

With β fixed, the convexity of the Langrangian is assured (Cover and Thomas, 2006). Differentiating, the unique minimum can be found at

$$\frac{\delta}{\delta p(y|x)} \Lambda(p(y|x), \beta) = 0 \tag{8}$$

2.1.6 Blahut-Arimoto algorithm

Finding the derivative and solving for $p(y|x)$ gives

$$p(y|x) = \frac{1}{Z} p(y) \exp[-\beta d(x,y)] \tag{9}$$

where Z is a normalising factor $Z = \sum_{y'} p(y') \exp[-\beta d(x,y')]$. The output distribution $p(y)$, can be determined from the conditionals by marginalising out X.

$$p(y) = \sum_x p(y|x)\, p(x) \tag{10}$$

The formulas for $p(y|x)$ and $p(y)$ given by Equations 9 and 10 are self-referential, but by initialising with any $\beta \in (0, \infty)$, any random set of conditional distributions for $p(y|x)$, and iterating between the two, the distributions converge on a solution. This elegant dual-minimisation iterative algorithm was invented simultaneously by two scientists (Blahut, 1972; Arimoto, 1972), so it is known as the *Blahut-Arimoto* algorithm.

Armed with the conditional distribution $p(y|x)$, the mutual information $I(X;Y)$ is straightforward to compute via Equation 3. The channel distribution $p(y|x)$ is optimal in the sense that for a required distortion, the mutual information $I(X;Y)$ is minimised. If the requirement is for no distortion, as $\beta \to \infty$, the mutual information $I(X;Y)$ will approach the minimum required for no distortion in this channel. β can be used to parameterise the trade-off between distortion and bandwidth for the channel, and this approach is used to similarly parameterise the trade-off between information and utility during the computation of policies based on minimising goal relevant information in Section 4.2.

2.1.7 Channel capacity

The channel capacity of the channel described in Section 2.1.4 above, is the maximum mutual information between X and Y.

$$\mathfrak{C} \coloneqq \max_{p(x)} I(X;Y) \tag{11}$$

The channel capacity, and corresponding source distribution $p(x)$ that maximises $I(X;Y)$, can be found using a similar approach to rate distortion, also formulated by Arimoto (1972) and Blahut (1972). The gothic $\mathfrak{C}$ is used to denote channel capacity here, to distinguish channel capacity from C for *graph centrality* in Section 2.5.

The flow of information in the perception-action loop, between the organism and its environment, can be modelled as a communication channel between states and actions. The information-theoretic tools outlined in this section will be applied to analysis of the perception-action loop in Section 2.4. The channel capacity of the actuation channel formulated via the treatment of the perception-action loop as a causal Bayesian network is *empowerment*, which is defined in Section 4.1.

2.2 Markov Decision Processes

2.2.1 Markov chains

In the early 20th century, the Russian mathematician Andrei Markov studied a type of *stochastic process* where the outcome of an event can influence the outcome of the next event. These sequences are known as *Markov chains*. A Markov chain is defined by a set of states $\mathcal{S} = \{s_1, s_2, ..., s_r\}$, and a transition matrix $\mathbf{P}_{ij} = p(s_j|s_i)$ where each element represents the probability of the process moving from state s_i to s_j. The process proceeds from an initial state, step-by-step, according to the transition probabilities. If $\mathbf{u}$ is a probability vector representing the initial distribution of states, then the probability of being in state s_i after t steps is the ith entry in the vector $\mathbf{u}^{(t)}$ where:

$$\mathbf{u}^{(t)} = \mathbf{u}\mathbf{P}^t \tag{12}$$

Markov chains are described as *regular* if for some power of $\mathbf{P}$ it has only positive entries. For regular Markov chains, the Fundamental Limit Theorem (Grinstead and Snell, 2006, p.447-451) states that there exists a unique matrix $\mathbf{W} = \lim_{t\to\infty} \mathbf{P}^t$ where each row of $\mathbf{W}$ is equal to a fixed probability row vector $\mathbf{w}$, which is straightforward to compute as it is the left eigenvector of $\mathbf{W}$. Furthermore, for any initial distribution of states $\mathbf{u}$, $\mathbf{u}\mathbf{P}^t \to \mathbf{w}$ as $t \to \infty$. This fixed distribution of states $\mathbf{w}$ in the limit of t is known as the *stationary distribution*.

2.2.2 MDP

A *Markov decision process* (MDP) is a discrete-time stochastic control process (Bellman, 1953; Howard, 1960). An MDP describes a model of an agent taking decisions, motivated by rewards at each time step t, and is defined by the tuple $\left(S, A, P_{s_t a_t}^{s_{t+1}}, R_{s_t}^{s_{t+1}}\right)$, where $S \in \mathcal{S}$ is a random variable modelling state, $A \in \mathcal{A}$ is a random variable modelling actions, $P_{s_t a_t}^{s_{t+1}} = p(s_{t+1}|s_t, a_t)$ is the *state-action transition function*, and $R_{s_t}^{s_{t+1}} \mapsto \mathbb{R}$ gives the *reward* r_t the agent receives for transitioning from state s_t to state s_{t+1}. The states of the agent over time form a first-order Markov chain, so the probability of transition from state s_t to s_{t+1} depends only on a_t, s_t and P, and not on any previous state $p(s_t|s_{t-1}, a_{t-1}) = p(s_t|s_1, a_1, ..., s_{t-1}, a_{t-1})$. Figure 1 depicts a MDP as a *causal influence diagram* (Howard and Matheson, 1984) with three states beginning at s_{t-1}.

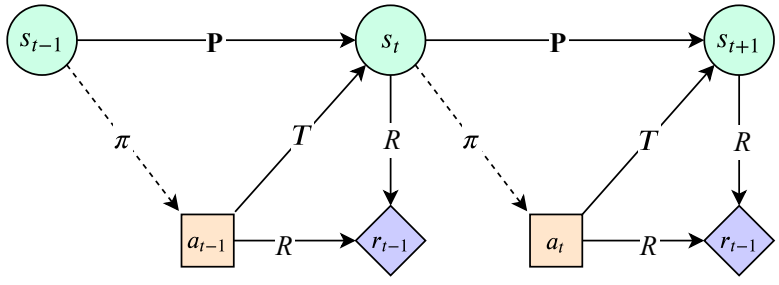


Figure 1: Markov Decision Process. The dashed arrow indicates the decision in the process, being the choice of action, made according to the stochastic policy π.

2.3 Reinforcement Learning

An agent acts in the world defined by a MDP to maximise the expected reward received, or minimise expected cost (rewards can be negative). The agent's strategy for choosing actions, depending on state, is implemented by a stochastic policy $\pi := p(a_t|s_t)$, which gives the probability of taking action a when in state s at time t. Beneficial actions are "reinforced" by reward, and the task of the agent is to learn a policy that maximises expected rewards. The policy and the reward function define a *value function* $V^\pi(s_t)$ that gives the time-unlimited expected sum of future rewards starting at state s_t.

$$V^\pi(s_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} ... = \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t R_{s_t}^{s_{t+1}}\Big] \tag{13}$$

The discount factor $\gamma \in [0, 1]$ controls the relative importance of short and long term rewards, and with a value of less than 1, has the practical effect of preventing an infinite sum for bounded rewards.

The value function can be written in recursive form as a *Bellman equation* (Bellman, 1953). This function is linear, so can be solved exactly using linear algebra methods, if the MDP is small enough for inverting the matrix to be tractable, or estimated to arbitrary precision by using the *value iteration* algorithm (Bellman, 1957).

$$V^\pi(s_t) = \sum_{a_t} \pi(a_t|s_t) \sum_{s_{t+1}} P_{s_t a_t}^{s_{t+1}} \big[R_{s_t}^{s_{t+1}} + \gamma V^\pi(s_{t+1})\big] \tag{14}$$

With the value function the agent can learn an optimal policy. Learning the policy with the transition and value functions already acquired is known as *model-based learning*, because the agent already has a model of the world, and just has to find an optimal policy.

$$\pi^* = \arg\max_\pi V^\pi(s) \tag{15}$$

If maximising utility is the only constraint, the optimisation problem in Equation 15 can be solved approximately and efficiently with a modified form of value iteration known as *policy iteration* (Russell and Norvig, 2009).

To evaluate the utility of actions under a policy, the *action-utility function* $Q^\pi(s_t, a_t)$ gives the time-unlimited expected sum of rewards after choosing an action a_t at state s_t.

$$Q^\pi(s_t, a_t) = \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a_t) \big[R_{s_t}^{s_{t+1}} + \gamma V^\pi(s_{t+1})\big] \tag{16}$$

In the Spiers lab Soho navigation experiment the participants were well trained on the map of Soho. The agent-based model of navigation employed in the present study correspondingly utilises a model-based approach where the transition function is known to the agent - it has a fully-observable model of the environment. The policy evaluation formula in Equation 16 is deployed across goals in Equation 28 in Section 4.2 to find a policy that satisfies a given utility constraint while minimising the required information in a dual-minimisation algorithm based on the Blahut-Arimoto algorithm in Equations 9 and 10.

2.4 Information in the perception-action loop

From the concept of living things as information processing entities, agent state, sensory observations (perception) and actions can be represented as a cyclical flow of information known as the perception-action loop (Polani et al., 2001).

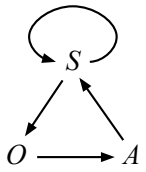


Figure 2: Perception-action loop. S represents agent and world state, O sensor observations about the state, and A actions taken by the agent.

Unrolled over time, the perception-action loop can be represented by a linked family of probability distributions known as Causal Bayesian Network (CBN) (Pearl, 1988). The agent receives information about the environment (observation) and acts on that information (action) to change state in a series of discrete time steps as shown in Figure 3 where the arrows represent the direction of causal influence. Random variables are only dependent on the other random variables from which they have an incoming arrow.

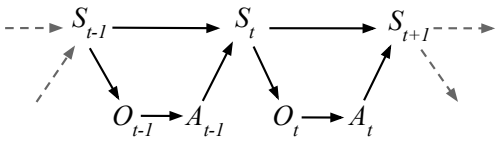


Figure 3: Perception-action loop unrolled into a Causal Bayesian Network. Random variables S_n, O_n, A_n model the state of the world, sensor observations and actions at time $t = n$

The CBN can be considered to model both the dynamics of the "flow of information" from the environment to the agent via its sensors, and the ability of the agent to detect changes it has made to the environment via its actuators (Klyubin et al., 2004). The mutual information $I(S;O)$ quantifies the average amount of information the agent acquires about the environment. The mutual information $I(O;A)$ quantifies how much of this information, on average, is used to select actions. If the sensors are perfect and the environment is fully-observable, the CBN can be reduced to a network of states and actions (Figure 4) which, with the introduction of rewards and a policy, becomes a Markov Decision Process (Figure 1). In this simplified model the mutual information between state and action $I(S;A)$, is the amount of information acquired and processed by the agent taking actions in the world.

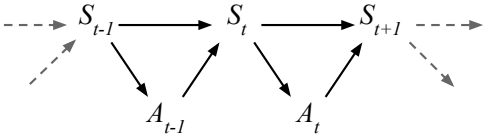


Figure 4: Simplified CBN for fully observable model.

2.4.1 Relevant Information

With the evolutionary driver to minimise information processing cost (information parsimony) in mind, the CBN induced by unrolling the perception-action loop can be considered as a communication channel, enabling the quantification of the amount of information an agent needs to take in from the world to make decisions. Considering the flow of information between states and actions as a communication channel, characterised by the agent decision making policy:

$$S_t \xrightarrow{\pi(a_t|s_t)} A_t$$

$I(S;A)$, the mutual information between the random variables modelling state and actions, is the amount of information that the agent needs to acquire from the environment to take action. $I(S;A)$ can also be thought of as the reduction in uncertainty about the action, given knowledge of the state.

$$I(S;A) = H(A) - H(A|S) \tag{17}$$

$I(S;A)$ varies according to the policy $\pi(a_t|s_t)$ selected by the agent. The agent is driven, by the information parsimony principle, to choose a policy with the lowest cost possible, while still operating successfully in the world. Finding the minimum average state-action information needed under a minimum utility constraint U_{min} formulates a new variational information problem. This quantity is known as *relevant information* (Polani et al., 2001, 2006) that quantifies the average information required to take action, and can be considered a measure of the cognitive burden of the agent (Polani, 2009).

$$\min_{\pi} I(S;A) \text{ subj. to } \mathbb{E}\big[Q^{\pi}(S,A)\big] \geq U_{min} \tag{18}$$

Equation 18 turns out to be formally equivalent to the rate distortion problem in Equation 5, with the distortion replaced by the negative utility and can be solved similarly by deploying a Lagrangian:

$$\Lambda(\pi,\beta) = I(S;A) - \beta\mathbb{E}\big[Q^{\pi}(S,A)|\pi\big] \tag{19}$$

Differentiating, so that an equation for the fixed point can be derived:

$$\frac{\delta}{\delta\pi}\Lambda(\pi,\beta) = p(s)\log\frac{\pi(a|s)}{p(a)} - \beta\, p(s)\, Q^{\pi}(s,A) \tag{20}$$

Setting $\frac{\delta}{\delta\pi}\Lambda(\pi,\beta) = 0$, and rearranging, gives the self-consistent solution

$$\pi(a|s) = \frac{1}{Z}p(a)\exp\left[\beta Q^{\pi}(s,a))\right] \tag{21}$$

The marginal distribution $p(a)$ is given by:

$$p(a) = \sum_{s} p(a|s)\, p(s) \tag{22}$$

Equations 21 and 22 are iterated to minimise $I(S;A)$ while achieving the utility given by $Q^{\pi}(s,a)$ as controlled by β. β controls the maximum utility, permitting the trade-off between utility and relevant information. While narrowing in on the minimum information, by minimising the difference between the action entropy and the action entropy given the state, the algorithm

also finds the policy $\pi(a|s)$ that satisfies the utility/information trade-off requirement controlled by β. As $\beta \to \infty$ the policy will tend to only include optimal actions. Where one action leads to a higher utility than others, under this regime the agent is certain to take that action. However if two actions have equal utility, the policy found will ensure the minimum relevant information is required, and the two actions will have equal probability.

As $\beta \to 0$ the optimisation minimises the average relevant information without regard to utility. The policy will tend to a random walk, where the agent requires no information at all to take action. Notably this does not necessarily result in zero expected utility, (or infinite cost).

The algorithm iterates between Equation 21 and 22, converging on an signal encoding that minimises the expected distortion for a given level of compression. A heavily compressed signal has the advantage of requiring less bandwidth, at the expense of the accuracy of the received signal when compared to the transmitted signal. In line with our consideration of intelligent agents as information processing entities, less bandwidth (relevant information) means less processing cost, which may be worth some loss of signal accuracy (utility).

Although β seems similar to a parameter controlling the balance of goal-directed and explorative behaviour used in artificial intelligence, it has a different role. As β is lowered the policies generated remain goal-directed (until $\beta = 0$), albeit at a cost of reducing utility, but with a commensurate reduction in the informational burden. The outcome of a lower β for the navigating agent is a policy where actions are shared across goals reducing the average information processing cost to the agent, at the expense of some increase in average distance.

An important difference in the formulation of relevant information, compared to the rate-distortion algorithm in Section 2.1, is that the utility function $Q^\pi(s, a)$, which determines the cost of the mapping, is not a fixed matrix, but depends on the policy. While iterating between the two minimisations, the utility function must be updated, by policy evaluation at each iteration, to incorporate the effect of the changing policy (Van Dijk, 2013).

The relevant information as described here is a global average. One could imagine dropping an agent into the world randomly, the relevant information would be the average state information required to take action. The formalism of an MDP does not require the unification of actions between states. It is perfectly permissible in the extreme case, to have a completely different set of actions in each state. It is clear from Equation 17 that the relevant information is lower bounded by the entropy of the action variable. In a world where no actions are shared between states, $H(A)$ will in general be much larger than where the same set of actions is available in every state. The effect of this on the relevant information is profound, with much more information required on average to take action. At the other end of the scale, if the same set of actions is reused at every state, the agent can reuse actions. For this to be possible the agent must be "placed" in the world – it must be "embodied". If the agent has four actions, North, East, South and West $\mathcal{A} \in \{N, E, S, W\}$ then it must have a sense of direction relative to the environment – it must know its allocentric orientation. Similarly if the agent has actions Forwards, Backwards, Left and Right, then it must know its egocentric orientation. The embodiment of the agent gives it the possibility to label actions to minimise the cognitive burden of operating in the environment (Polani, 2011). To elaborate an example pertinent to this thesis, one can imagine a nightmarish version of London with randomised action labels, where the outcome of going "North", "South", "East" or "West" is different at each junction. When advising a tourist asking directions, instead of saying "keep going East down Oxford street until you reach Tottenham Court Road" one would have to provide a list of actions as long as there are junctions en-route. The total information required about the route is much higher in this "twisted world" scenario.

The model described in Section 3, used for all experiments described in this thesis, defines that the same set of actions are available to the agent in all states.

2.5 Graph Centrality

A graph $\mathbf{G} = (V, E)$ consists of a set of vertices $V = \{v_1, v_2, ..., v_n\}, n = |V|$ and set of edges that connect pairs of vertices $E = \{e_1, e_2, ..., e_m\}, e = (u, v), u \in V, v \in V, m = |E|$. The graphs used in this thesis are *directed*, meaning that each edge is an ordered pair of vertices, connecting the two vertices in one direction only. In a *digraph* each pair of vertices may have an edge in one or both directions. Furthermore all graphs used herein are *connected*, meaning that there is at least one path of edges connecting all pairs of vertices.

2.5.1 Degree Centrality

The *degree* of a vertex is defined as the number of other vertices to which it is connected (by outgoing edges as the graphs used here are directed graphs). Degree *Centrality* is sometimes defined as normalised degree, but following the convention in Javadi et al. (2017), the degree centrality of a vertex is defined here as simply the degree.

$$C_D(v) = \deg(v) \tag{23}$$

2.5.2 Closeness Centrality

Closeness centrality of a vertex is the reciprocal of the total shortest path distances to all other vertices (Brandes and Erlebach, 1998). $l(u, v)$ between two vertices u and v is defined as the length of the shortest path (graph geodesic) from u to v.

$$C_C(v) = \frac{1}{\sum_{u \in V} l(u, v)} \tag{24}$$

2.5.3 Betweenness Centrality

Betweenness centrality of a vertex is the fraction of all the geodesics (shortest paths) between all pairs of vertices that pass through the vertex.

$$C_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{25}$$

where σ_{st} is the number of geodesics between vertices s and t, and $\sigma_{st}(v)$ is the number of geodesics between s and t that pass through v (Brandes and Erlebach, 1998).

2.6 Software tools

All algorithms and computations were developed in Python using NumPy[2], SciPy[3], and the high precision library GMPY[4]. For probability and information-theoretic computations I used DIT[5]

[2] https://pypi.org/project/numpy/
[3] https://pypi.org/project/scipy/
[4] https://pypi.org/project/gmpy2/
[5] https://pypi.org/project/dit/

(James et al., 2018), and NetworkX[6] (Hagberg et al., 2008) to represent the graphs and compute primal graph centrality measures.

[6]https://pypi.org/project/networkx/

3 Experimental Model

3.1 Information processing agents

Animals operate in the world by acquiring information through the senses. Sensory information is processed to enable advantageous behaviour. A simple example is a nematode worm sensing and moving along a chemical gradient signifying the direction of a food source.

To study navigational behaviour I employ a discrete probabilistic model that is commonly used in the study of artificial intelligence. The world consists of a set of discrete states representing locations which can be mapped onto a planar graph (see Figures 6, 7 and 9). An autonomous agent can take action by moving from one state to another. The agent tries to a reach a goal by taking in sensory information about the world and acting accordingly. For simplicity, the agent is assumed to have perfect sensors - the world is fully observable and it is not necessary to describe the agent's belief about state as distinct from the actual state. The agent acting in the world takes place as a series of time steps, each associated with random variables describing the agent's state $S_t \in \mathcal{S}$ and actions $A_t \in \mathcal{A}$. A time-independent random variable models the location of the goal $G \in \mathcal{G}$. As described in Section 2.4 the cycle of processing sensory information at each time step and acting accordingly is known as the *perception-action loop*. The perception-action loop of the agent is modelled by a Causal Bayesian Network (CBN) (Pearl, 1988). The agent receives information about the environment (state) and acts on that information (action) in a series of discrete time steps. The CBN can be considered to show both the dynamics of the "flow of information" from the environment to the agent via its sensors, and the ability of the agent to detect changes it has made to the environment via its actuators (Klyubin et al., 2004). In this model agent sensors are perfect – the environment is fully-observable, so the CBN consists of a network of states, actions and the goal as depicted in Figure 5. Because the state at any given time is only dependent on the previous state, the CBN unrolled in time forms a Markov Chain, and the behaviour of the agent may be analysed through the lens of reinforcement learning (see Section 3.3).

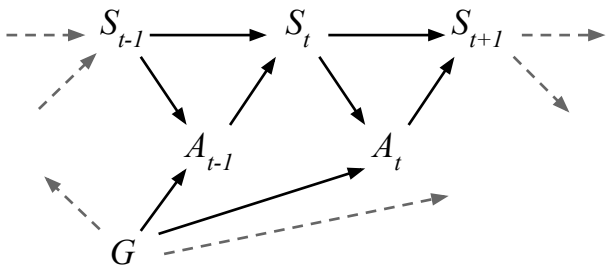


Figure 5: Perception-action loop unrolled into a Causal Bayesian Network. S, A and G are random variables modelling states, actions and goals respectively. S_t is the state of the world at time t, A_t is the action taken by the agent at time t that leads to the new world state in the subsequent time step S_{t+1}. The arrows show the direction of causality. The current state is dependent on the previous state and the previous action taken at that state. The current action is dependent on the current state and the goal.

3.2 Graph representation

The study of human navigational behaviour in cities over four decades employs a range of graph-theoretic measures (Hillier and Hanson, 1984), to predict macroscopic pedestrian and traffic flow.

A graph can conveniently represent the junctions (vertices) and streets (edges) of a street network, allowing the graph-theoretic centrality of street networks to be measured. The same graph can be used to represent the states (vertices) and actions (edges) of the simple discrete probabilistic model of an autonomous agent as described above.

Simple grid worlds are commonly used in the study of artificial intelligence, and specifically in the prior work on *empowerment* (Klyubin et al., 2004, 2005*a*; Salge et al., 2014*b*; Clements and Polani, 2017) and *relevant goal information* (Van Dijk et al., 2010; Van Dijk and Polani, 2011, 2012) that form the foundation for this thesis. Although in the literature grid worlds are naturally diagrammed as grids, I will show them as graph diagrams for continuity with the graph diagrams of the street networks, which naturally resemble cartographic maps.

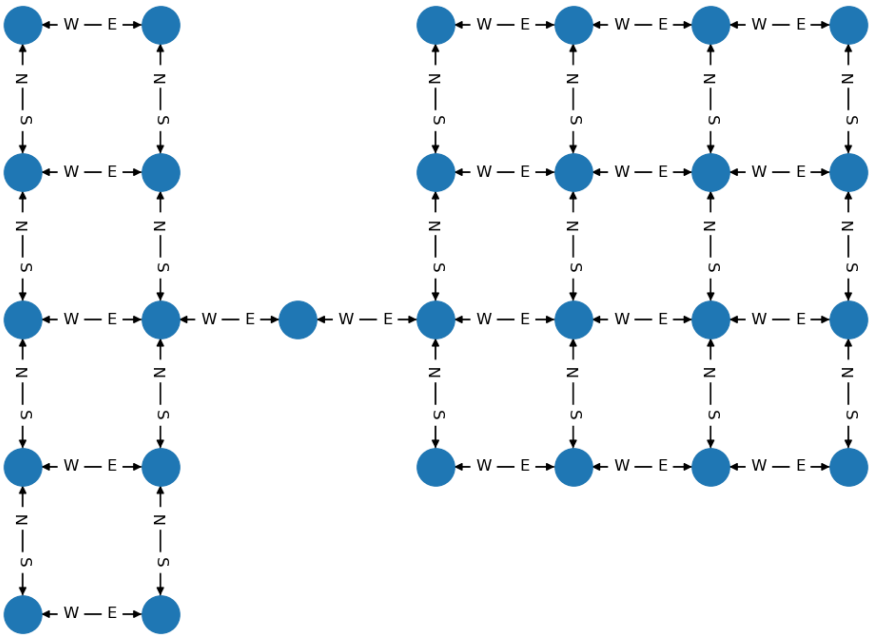


Figure 6: Example of a small grid world. The agent has five possible actions in any state $\mathcal{A} = \{N, E, S, W, wait\}$. All rewards are -1.

There are two types of world employed in the experiments, *grid worlds* and *street networks*. Each world is represented by a connected directed planar multigraph $\mathbf{G} = (V, E)$ with vertices $V = \{s_1, s_2, ..., s_n\}, n = |V|$ representing all possible agent states, and edges representing all possible state-action transitions $E = \{e_1, e_2, ..., e_m\}, e = (s, s'), m = |E|$. Each edge is associated with one of the actions from the set of actions available to the agent, $a \in \mathcal{A}$. Figure 6 shows an example small grid with edges labeled with actions. Figure 7 shows a graph of a small section of the Soho street network with edges labeled with actions.

The graph partially represents a MDP (see Section 3.3), with the set of vertices being the alphabet of states $\mathcal{S} = V$, and the set of edges defining the state-action transition function. If an edge connects two vertices, when the agent takes the action represented by the edge, it will transition between the states. If no edge exists between the states corresponding to the action then taking the action results in no change of state. Formally, where a is the action connecting s_i to s_j and $s_i \neq s_j$:

$$\Pr\left(S' = s_j | , S = s_i, A = a)\right) = \begin{cases} 1, & \text{if } (s_i, s_j) \in E \\ 0, & \text{otherwise} \end{cases}$$

and

$$\Pr\left(S' = s_i | , S = s_i, A = a)\right) = \begin{cases} 0, & \text{if } (s_i, s_j) \in E \\ 1, & \text{otherwise} \end{cases}$$

The choice of a directed graph representation permits a general transition model where the probability or cost of transitioning between two states may be different in either direction (e.g. one-way street, or restricted flow in one direction). For simplicity, and since pedestrian traffic typically flows both ways in all streets, all the graphs used in the experiments below have both "in" and "out" edges between each pair of vertices, and the cost of transitioning in either direction is the same. Graph representations of MDPs can also provide performance enhancements for estimating policies e.g. (Cheng and Chen, 2013). Exploiting such enhancements presents an opportunity for further work.

3.2.1 Street Networks

For street networks, the planar graph representation has the advantage of resembling a cartographic map. Street network graphs shown in figures in this text are shown with variable length edges representing the length of the street segment in the map. These lengths are for diagrammatic purposes only, the reward function just gives -1 for each transition, the same as for grid worlds. The vertices represent junctions and the edges represent the action of walking along a street between two junctions. All streets were assumed to be navigable in either direction by pedestrians, so all vertices are connected to neighbours by a pair of edges, one for each direction. The cardinality of the set of actions is determined by the maximum degree of the vertices of the graph $|\mathcal{A}| = \Delta\mathbf{G} + 1$. The choice of action for an edge is decided by bearing. Analysis of the distribution of bearings in Soho street network with a polar histogram (Figure 8) shows a strong clustering of bearings in four directions approximately aligned with the intermediate directions (NE, SE, SW, NW). The set of all bearings were analysed with the *k-means* algorithm, to find bearings that represented $|\mathcal{A}|$ cluster centroids. Actions were allocated to edges in turn, with the edges ordered by angular distance from cluster centroid bearings. The junctions in the street maps used in the experiments had a maximum of 6 connected streets (at radius 500m from centre), represented by a graph with a maximum degree of 6, with 7 actions in the set of actions available to the agent in each state, once the "wait" action is included. Section 3.2.2 describes the process of generating graphs from cartographic data.

3.2.2 Generating graphs from maps

The street network graphs were generated from the Space Syntax map file used for Javadi et al. (2017), kindly supplied by Joao da Silva Pinelo. The map includes data for 36,319 street segments covering all of central London. Computing relevant goal information (see Section 4.2) would be intractable on such a large state space, so the set of street segments included in the graphs was limited by distance from the centre point of the street network used in the study. Graphs were generated from the set of street segments, where both ends of the segment are within the chosen radius for radii $r \in \{300, 400, 500, 600, 700\}$ metres. The graph at 300m radius was the smallest

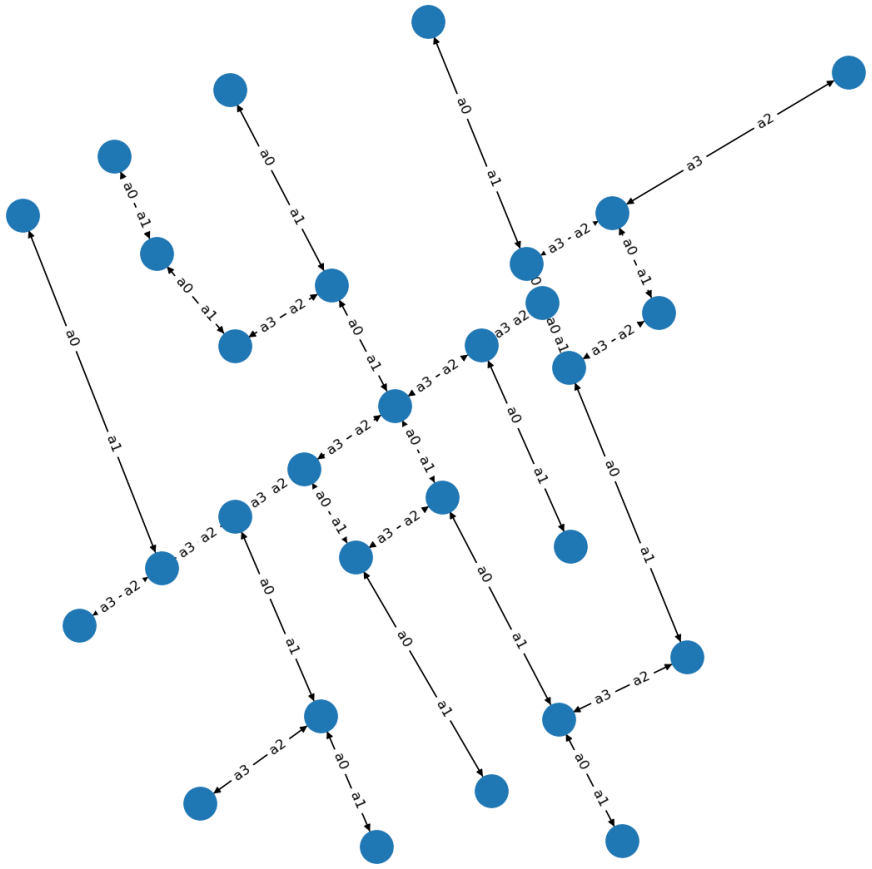


Figure 7: Section of the Soho street network graph showing action labels, 100m radius from centre. The agent has five possible actions, $\mathcal{A} = \{a1, a2, a3, a3, wait\}$, because no junction connects more than four streets. Actions have been allocated according to which cluster of bearings they fall into. Vertices are junctions, edges are streets.

radius where no junctions of interest were excluded from the graph, the graph at 500m radius was found to have the best correlation with the space syntax centrality measures (see Section 5).

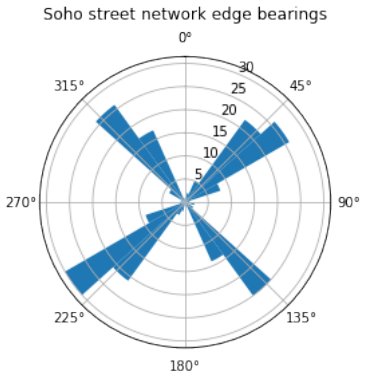


Figure 8: Polar histogram, generated from the Space Syntax map data, showing distribution of bearings in the Soho street network amongst streets with both junctions within 500m from centre.

Figure 9: Blue and red points show the graph of radius 500m generated from Space Syntax map data overlaid on a Google map of the area around Soho, London. Red points are junctions in the network used in Javadi et al. (2017).

3.3 A family of Markov Decision Processes

The Causal Bayesian Network in Figure 5 can be viewed as a series of decisions in a fully observable stochastic environment, where the next action is only dependent current state, not any previous states or actions, via the stochastic agent policy and the deterministic transition function. Unrolled in time the CBN represents a family of Markov Decision Processes (see Section 2.2), each one describing how the agent seeking a goal $g \in \mathcal{G}$ in state $s_t \in \mathcal{S}$ at time t takes action $a_t \in \mathcal{A}$, thus transitioning to state $s_{t+1} \in \mathcal{S}$. The MDPs are defined by the tuple $\left(S, A, P_{s_t a_t}^{s_{t+1}}, R_{s_t g}^{s_{t+1}}\right)$, where S is a random variable modelling state, A is a random variable modelling actions, $P_{s_t a_t}^{s_{t+1}} = p(s_{t+1}|s_t, a_t)$ is the *state-action transition function* – in this model a deterministic mapping of states and actions defined by the graph. $R_{s_t g}^{s_{t+1}} \mapsto \mathbb{R}$ is the *reward* the agent receives for transitioning from state s_t to state s_{t+1} while seeking goal g. The reward models the agent's motivation to reach the goal as easily and quickly as possible. Rewards are negative and represent the cost of acting in the world. Some actions result in a change of state, which incurs a physical cost of movement, some actions do not result in a change of state, but nonetheless incur a cost. Although it might seem simpler to restrict the set of available actions in each state, A_t models the agent's choice of action from the normal set available to it. The agent "could" walk into a wall, but informed by the negative reward, it follows a policy where this action is unlikely.

At each time step the agent chooses actions based on the policy – a time-invariant probability distribution of actions given state $\pi = p(a|s)$. The deterministic transition function in this model simplifies computation without loss of expressivity of the features of interest, however the stochastic policy is crucial to the model. Acting optimally in all situations comes with a cost, a stochastic policy allows the agent to save informational bandwidth and/or storage by accepting some loss of optimality in certain situations. Considering the navigation scenario, at some junctions it doesn't matter much, in terms of distance travelled to most goals, whether we go left or right. With a stochastic policy the agent can assign a similar probability to these actions. As we shall see in Section 4.2 this is where the power of information theory can be brought to bear. The agent has previously acquired the policy through knowledge of the environment and the goal, and the policy is not updated in the course of an experiment. The agent taking decisions according to the policy, combined with the outcome of the state-action transition function, forms a Markov chain of states defined by the state-transition matrix $\mathbf{P}$ with elements set as $p_{ij} = \Pr(S_{t+1}{=}j|S_t{=}i) = \sum_{a_t} \pi(a_t|s_t)\, P_{s_t a_t}^{s_{t+1}}$

3.4 Utility and policy evaluation

The model utilises standard reinforcement learning formalism to evaluate agent policies (see Section 2.3). The expected reward for executing the policy in a given state at time t is the *value function* $V_g^\pi(s_t)$, defined as the expected sum of rewards at state s, for a time unlimited sequence of actions chosen according to the policy $\pi(a_t|s_t, g)$ specified by the goal g . There is no need to limit the horizon of the agent with a discount factor because in this model MDPs are always finite – the goal is absorbing and the agent never has a policy that actively avoids the goal, so it will always arrive at the goal eventually.

$$V_g^\pi(s_t) = \mathbb{E}\big[r_t + r_{t+1} + r_{t+2} + ...\big] \tag{26}$$

The value function can be expressed as a recursive Bellman (1953) type equation:

$$V_g^\pi(s_t) = \sum_{a_t \in \mathcal{A}} \pi(a_t|s_t, g) \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a_t)\big[R_{s_t g}^{s_{t+1}} + V_g^\pi(s_{t+1})\big] \tag{27}$$

To evaluate the utility of actions at a particular state under a policy the *action-utility function* $Q_g^\pi(s_t, a_t)$ gives the time-unlimited expected sum of rewards after choosing an action a_t at state s_t.

$$Q_g^\pi(s_t, a_t) = \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a_t)\big[R_{s_t g}^{s_{t+1}} + V_g^\pi(s_{t+1})\big] \tag{28}$$

$V_g^\pi(s_t)$ can be estimated to arbitrary precision using the well-known *value iteration* algorithm (Russell and Norvig, 2009), and this is the approach taken by Polani et al. (2006) and Van Dijk et al. (2010).

3.4.1 Policy evaluation via the fundamental matrix

The Markov chain formed by the transition matrix for a goal directed agent, where the goal is an absorbing state, is an absorbing Markov chain. The mean time to absorption can be found via the *fundamental matrix* (Grinstead and Snell, 2006). For an MDP where the reward is the same for each transition (in this model $r=-1$), the mean time to absorption gives the expected value of the policy $V_g^\pi(s_t)$.

First I arrange the transition matrix $\mathbf{P}$ such that the goal (absorbing state) is in the last row and column of the matrix, with the transient (non-absorbing) states occupying the rest of the row and column.

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{0} & 1 \end{array}\right)$$

$\mathbf{b}$ is a column vector representing the probability of transitioning into the goal at each non-goal state $\mathbf{b} = \Pr(S_{t+1}{=}g|S_t)$, $\mathbf{0}$ is a zero row vector representing the probability of transitioning out of the goal into any other state $\mathbf{0} = \Pr(S_{t+1}|S_t{=}g)$. The matrix $\mathbf{A}$ represents the probability of transitioning between any pair of transient states $\mathbf{0} = \Pr(S_{t+1}|S_t{\neq}g)$. The matrix $\mathbf{N} = (\mathbf{I} - \mathbf{A})^{-1}$ is known as the *fundamental matrix* for $\mathbf{P}$. The entries n_{ij} of $\mathbf{N}$ give the expected number of times the process is in s_j if started in s_i. The expected time to absorption (expected time to reach the goal) is given by $\mathbf{t}_g = \mathbf{Nc}$, where $\mathbf{c}$ is a column vector with all entries equal to 1, and $\mathbf{t}_g$ is a vector of mean time to absorption for each state except the goal (Grinstead and Snell, 2006). Section 4.2.2 discusses some performance implications of using this approach as opposed to value iteration.

4 Information-theoretic measures

Underlying the hypothesis is the proposal that an information-theoretical analysis of an agent-based model of human navigation can explain the ability of a graph-theoretic analysis of the street network to predict human navigational behaviour. This section outlines the formalism and computation of the information-theoretic measures empowerment, relevant goal information and relevant goal information uptake, as applied to the discrete probabilistic model described in Section 3.

4.1 Empowerment

The cycle of intake of sensory information and taking action based on that information is known as the perception-action loop, which can be considered to be a communication channel (see Section 2.4. Empowerment ($\mathfrak{E}$) is defined as the *channel capacity* ($\mathfrak{C}$) (Shannon, 1948) of the actuation channel of the agent, and is formalised as the maximal possible information flow between the actions of the agent and the effect of those actions some time later. Empowerment can be interpreted as the potential the agent has to influence the environment. Empowerment can be computed for a given number of cycles into the future, which is referred to in the literature as *n-step* empowerment (Klyubin et al., 2005*b*). Since I use the binary logarithm, empowerment will be measured in *bits*.

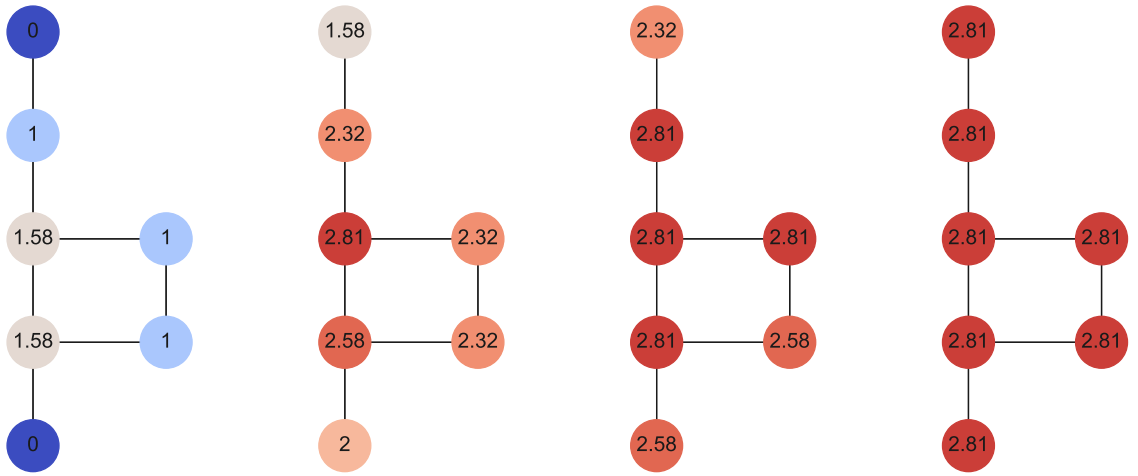
$$\mathfrak{E}(s_t) \coloneqq \mathfrak{C}(s_t) = \max_{p(a_t^n|s_t)} \sum_{\mathcal{A}^n,\mathcal{S}} p(s_{t+n}|a_t^n, s_t)\, p(a_t^n|s_t)\, \log \frac{p(s_{t+n}|a_t^n, s_t)}{\sum_{\mathcal{A}^n} p(s_{t+n}|a_t^n, s_t)\, p(a_t^n|s_t)} \tag{29}$$

In this deterministic transition, fully-observable model the computation simplifies to the log of the number of reachable states within n steps.

$$\mathfrak{E} = \log\, |\mathcal{S}_n| \tag{30}$$

Empowerment has been shown to be useful in a range of settings as a task-less utility function for a maximum empowerment policy, where, in the absence of information about the goal, an agent can choose the action that maximises the future opportunities that are available within the time window (Klyubin et al., 2008). The variable range of n-step empowerment means it can utilise from local to global topology, depending on the ratio of the number of steps to the size of the environment.

Figure 10 shows empowerment calculated for the states in a very small graph world. An agent navigating this world has 5 actions available in every state $\mathcal{A} = \{N, E, S, W, \text{wait}\}$. The "wait" action, and any action without a corresponding edge in the graph, result in no change of state $s_{t+1} = s_t$. In dynamic environments unchanging state can lead to higher empowerment but in a static environment such as this model, with no change in state there is no empowerment. In 1 step, the agent can reach each connected state. $\mathfrak{E}_{1-step}(s) = \log \deg v_s$ where s is a state and v_s is the graph vertex representing s, hence the maximum empowerment is set by the maximum degree of the vertices of the graph ($\Delta\mathbf{G}$) $\max \mathfrak{E}_{1-step}(s) = \log \Delta\mathbf{G}$. Figure 10a shows the 1-step empowerment for a small graph. The empowerment for a "dead end" state with only one neighbour is zero. At 2 steps for the small world in Figure 10b, empowerment uncovers some aspects of the topology. There are two states with a degree of 3 but one has a higher empowerment because one more state is reachable with 2 steps. Similarly at 3 steps in Figure 10c, the two states to the right of the figure are differentiated. With 4 steps of look-ahead in Figure 10d, all the states are reachable, so the agent has maximum empowerment in every state.



(a) $\mathfrak{E}_{n_{steps}=1}$ (b) $\mathfrak{E}_{n_{steps}=2}$ (c) $\mathfrak{E}_{n_{steps}=3}$ (d) $\mathfrak{E}_{n_{steps}=4}$

Figure 10: Empowerment $\mathfrak{E}_{n_{steps}\in[1,4]}$ in *bits* for a small grid world. 1 step empowerment is the log of vertex degree. 4 step empowerment is the same for all states ($\log_2 7$) because any state can be reached in 4 steps from any other state. 2 and 3 step empowerment encode certain aspects of the topology of the world; at 2 steps the left central node has the highest empowerment. An agent could maximise opportunities in the absence of a clear goal by being at that location.

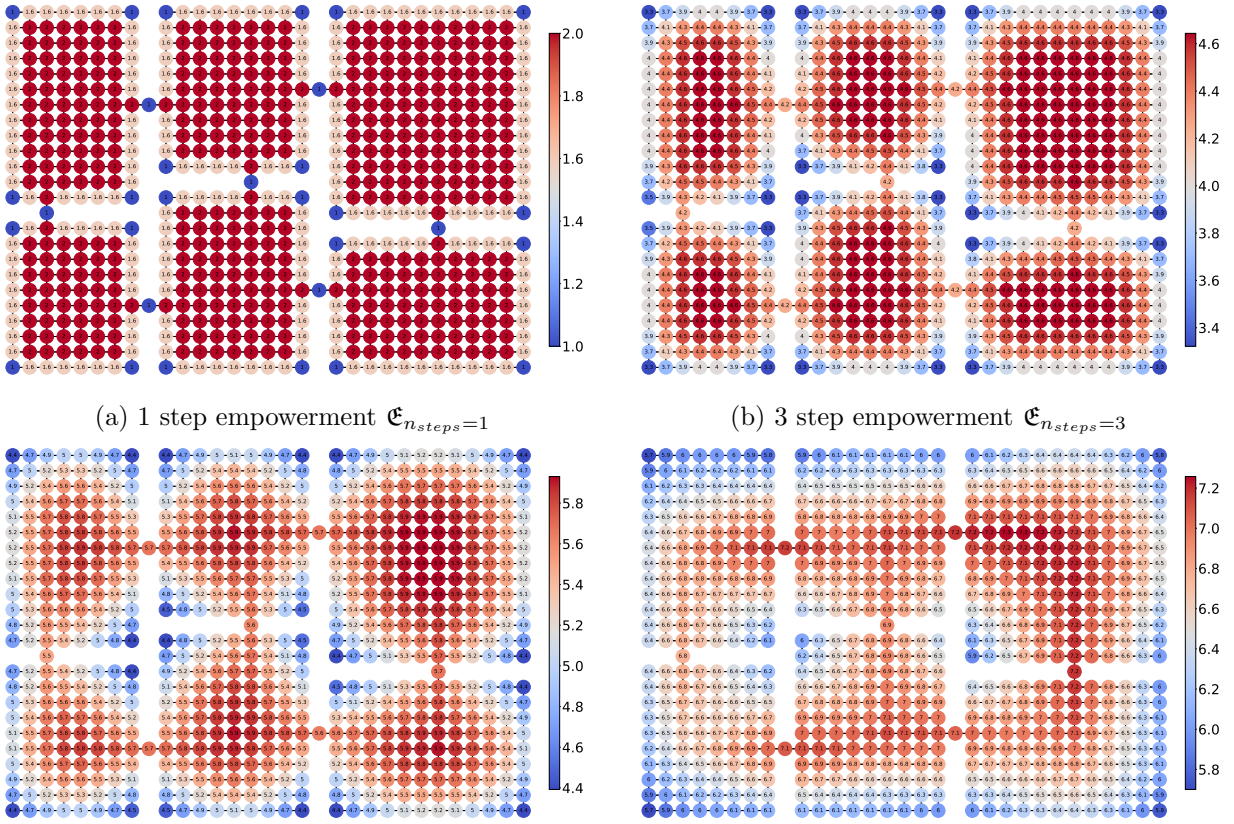
Three important factors have been identified to take into account for the choice of a suitable number of steps for fruitful analysis of graph worlds are:

1. At one step empowerment is proportional to vertex degree.
2. With a small number of steps relative to the diameter of the graph, empowerment encodes aspects of topology.
3. As the number of steps approaches the diameter of the graph, the spread of empowerment values amongst the states becomes less varied

The frequency distribution of empowerment values from 1 to 9 steps in Figure 13 illustrates this effect for the street network graph.

4.1.1 Empowerment for the six room grid world

Figure 11 shows empowerment at 1, 3, 5 and 9 steps for a six room grid world, familiar from the literature (Van Dijk et al., 2010). As the number of steps increases, extending the horizon of the agent, empowerment uncovers the importance of the doorways between the rooms, in providing an opportunity for more possible future states. For a comparison between empowerment and centrality for the six room grid world, please see Section 5.



(a) 1 step empowerment $\mathfrak{E}_{n_{steps}=1}$

(b) 3 step empowerment $\mathfrak{E}_{n_{steps}=3}$

(c) 5 step empowerment $\mathfrak{E}_{n_{steps}=5}$

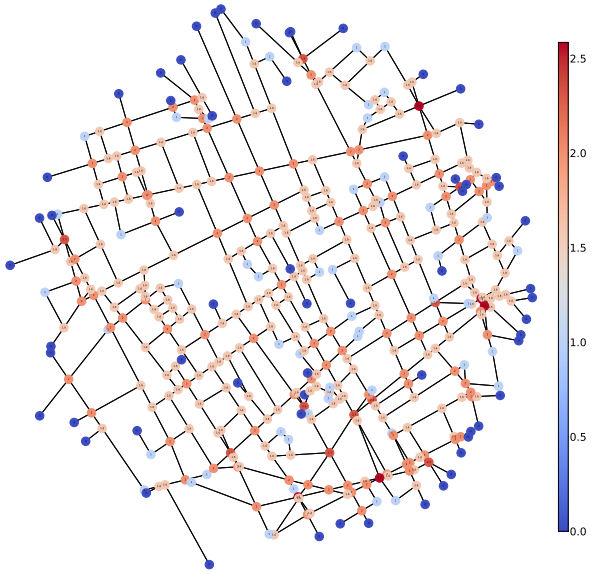
(d) 9 step empowerment $\mathfrak{E}_{n_{steps}=9}$

Figure 11: Empowerment $\mathfrak{E}_{n_{steps}\in\{1,3,5,9\}}$ for the six room grid world. Colour ranges are specific to each graph. At 1 step the measurement of empowerment is discrete and coarse, because empowerment is proportional to degree, with more steps empowerment elucidates network topology. At 3 and 5 steps the central regions of the rooms, the corridors and states close to the corridors allow access to other rooms so the empowerment is high. At 10 steps the relative size of the rooms has a strong influence on the value of empowerment.

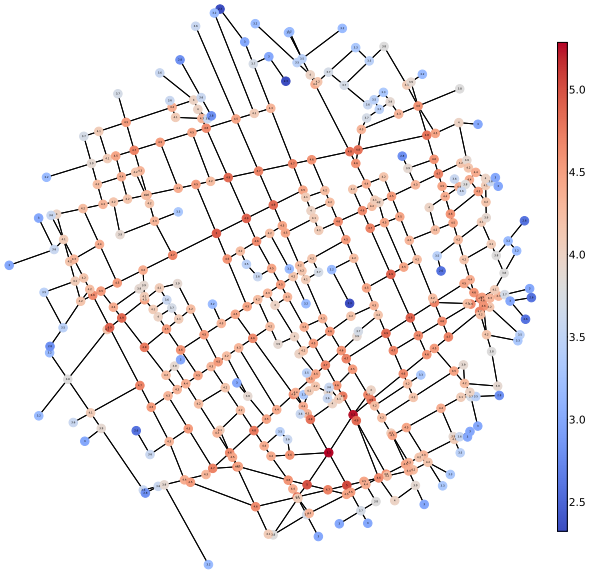
4.1.2 Empowerment for the street network graph

To illustrate the effect of increasing the number of steps in a street network, Figure 12 shows the changing distribution of empowerment values with increasing number of steps in a graph of the Soho street network including streets within a 500m radius of the centre. At 9 steps the ability of empowerment to distinguish between states is reduced as most states have a similarly high value. The edge effect means that states on the periphery of the graph will have a lower empowerment than the same states in a graph constructed with a larger radius. More states will be affected by this peripheral effect with a higher number of steps. Consequently the accurate computation of n-step empowerment requires that the junctions of interest are more than n steps from the edge of the map. Any closer and the values of empowerment would be artificially constrained by the choice of map radius, because the edge of the map could be reached within the n-steps. Up to 4 steps, the values of empowerment for the vertices (junctions) of interest do not change for radii of 500m and higher. At 5 or more steps, the value of empowerment is reduced, for junctions of interest closer to the edge of the map.

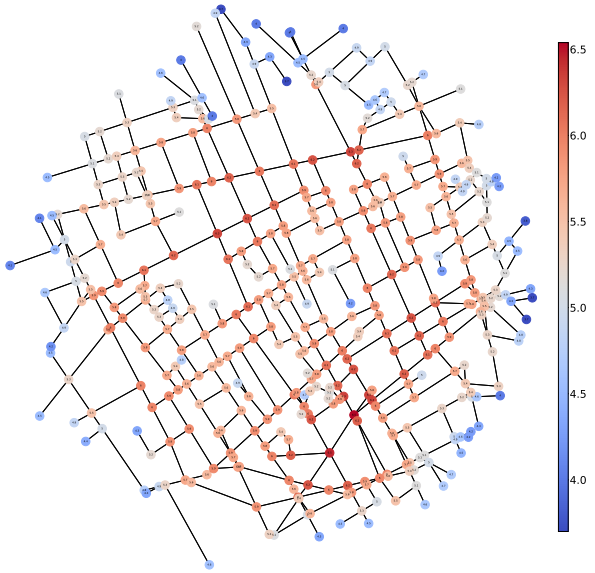
Figure 13 shows the frequency distribution of empowerment values in the 500m radius graph, from 1 to 9 steps. At 6 steps the maximum empowerment value is shared by 25 states, limiting the effectiveness of the measure to distinguish between states in the network.



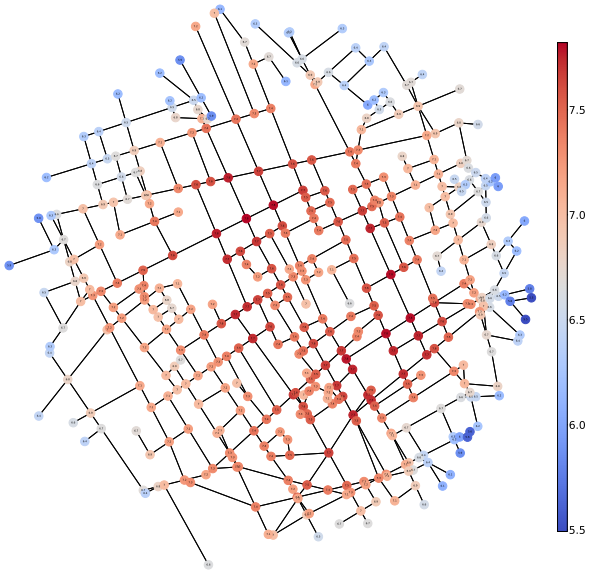
(a) Empowerment $n_{steps} = 1$



(b) Empowerment $n_{steps} = 3$



(c) Empowerment $n_{steps} = 5$



(d) Empowerment $n_{steps} = 9$

Figure 12: Empowerment $n_{steps} \in \{1, 3, 5, 9\}$ for all vertices in the Soho street network at 500m radius. By 9 steps many vertices have a relatively high value, and some of the vertices of interest will be less than 9 steps from the periphery, affecting their empowerment relative to more central states. Colour ranges are specific to each graph.

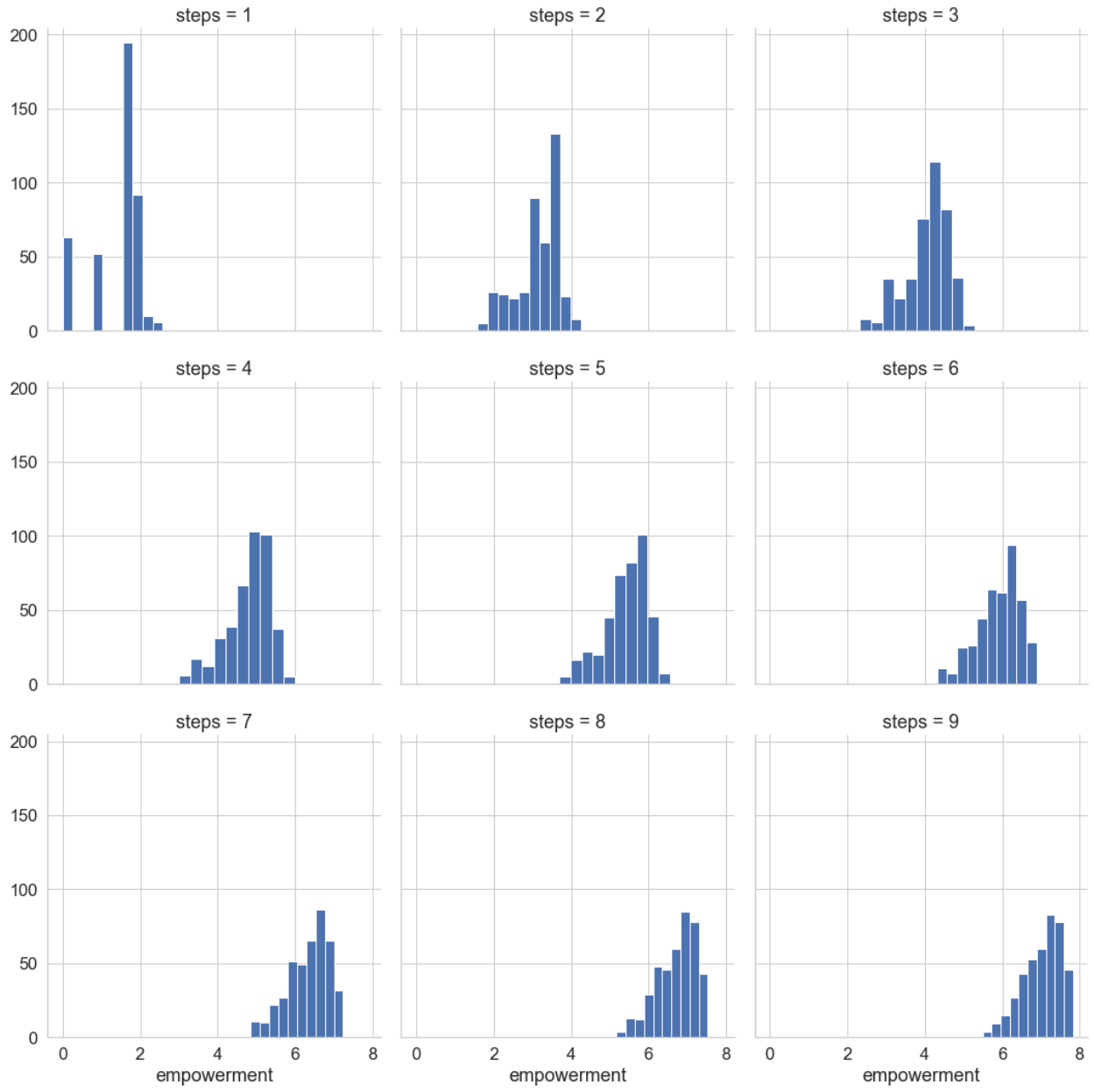


Figure 13: Frequency distribution of empowerment values from 1 to 9 steps for the Soho street network with a radius of 500m. At 6 steps and above there are a large number of states (25+) with maximum empowerment which limits the ability of the agent to differentiate between states. The frequency distribution of empowerment could be seen as an empowerment based signature of a street network, characterising the network for comparison with others.

4.2 Relevant Goal Information

Relevant goal information (RGI) is the minimum average amount of information about the goal required by the agent to select an action under a utility constraint (e.g optimal) and is defined as the *conditional mutual information* between the variables modelling the goal and action $I(G; A_t|s_t)$, given the state (Van Dijk et al., 2010). RGI can also be thought of as the reduction of the uncertainty about the choice of action provided by knowledge of the goal.

From the definition of mutual information(Shannon, 1948):

$$I(G;A_t|S_t) = H(A_t|S_t) - H(A_t|G, S_t) \tag{31}$$

To model the trade-off between physical and cognitive effort when navigating, the agent can choose a policy that minimises the average information required about the goal to achieve a certain utility, or a policy that maximises the utility for a certain amount of information about the goal. Choosing the former Van Dijk et al. (2010) define the following minimisation problem, the solution of which is the minimum information required, on average, about the goal, to choose an action that achieves a certain expected utility. This quantity is known as *Relevant Goal Information* (RGI).

$$RGI \coloneqq \min_{\pi(a_t|s_t,g)} I(G; A_t|S_t) \text{ subj. to } \mathbb{E}\big[Q^{\pi}_G(S_t, A_t)\big] \geq U_{min} \tag{32}$$

RGI as defined in Equation 32 is a global quantity averaged across all states and goals. The measure used to compare with graph centrality is $RGI(s_t)$, defined in Section 4.2.3, which is the relevant goal information for a particular state, averaged across all goals. U_{min} is a lower bound for the expected value of the utility function averaged across all goals subject to the policy π. To solve the minimisation problem in Equation 32, a similar reformulation to relevant information employing a Langrangian is used (see Section 2.4.1). The relevant goal information formulation and dual-minimisation outlined here are all due to (Van Dijk, 2013, Chapter 6), reproduced here for clarity. Once again β is a Lagrange multiplier constraining the utility.

$$\min_{\pi(a_t|s_t,g)} \big[I(G; A_t|S_t) - \beta\mathbb{E}[Q^{\pi}_G(S_t, A_t)]\big] \tag{33}$$

The solution of this fixed point equation is the relevant goal information. The algorithm also outputs a policy per goal represented by the conditional distribution $\pi(a_t|s_t, g)$ which minimises the RGI, within the expected utility constrained by β.

The Langrangian is given as:

$$\Lambda\big(\pi(a_t|s_t, g), \beta\big) = I(G; A_t|S_t) - \beta\mathbb{E}\big[Q^{\pi}_G(S_t, A_t)\big] \tag{34}$$

Taking the partial derivative with respect to $p(a_t|s_t, g)$

$$\frac{\partial}{\partial p(a_t|s_t, g)}\Lambda\big(\pi(a_t|s_t, g), \beta\big) = p(s_t, g)\, \log \frac{p(a_t|s_t, g)}{p(a_t|s_t)} - p(s_t, g)\, \beta Q^{\pi}_g(s_t, a_t) \tag{35}$$

Equating this derivative to zero and rearranging gives the self-consistent solution familiar from relevant information and so also rate-distortion:

$$p(a_t|s_t, g) = \frac{1}{Z}\pi(a_t|s_t) \exp\big[\beta Q^{\pi}_g(s_t, a_t)\big] \tag{36}$$

where Z is a normalising factor $Z = \sum_{a_t} p(a_t|s_t)\, \exp\left[\beta Q^{\pi}_g(s_t, a_t)\right]$. With a weighted sum over all goals the policy is obtained from the conditional distributions:

$$p(a_t|s_t) = \sum_g p(g)\, p(a_t|s_t, g) \tag{37}$$

4.2.1 Finding a minimum RGI policy to compute RGI

This section describes how to compute an agent policy that achieves an expected utility, while minimising the relevant goal information. With that policy in hand, the average RGI required per state can be computed for a given environment, such as the Soho street network, thus providing a state by state measure to compare with graph centrality.

With Equations 36 and 37 in the toolbox, the RGI minimisation problem in Equation 33 can be solved using the same method as relevant information as described in Section 2.4.1 by iterating between the two with a Blahut-Arimoto style dual-minimisation. The utility is constrained by the value of the Lagrange multiplier β. The policy must be evaluated at each iteration, because the expected utility depends on the policy. Two approaches to policy evaluation were investigated, approximation by value iteration and exact computation by linear algebra – both are outlined in Section 4.2.2

The algorithm is initialised with a uniform policy $\pi^0(a_t|s_t)$ where $\forall a; a{\in}A, \pi^0(a_t|s_t){=}\frac{1}{|A|}$. The uniform policy represents a *random-walk* by the agent which is clearly sub-optimal in terms of utility, but requires no goal information at all. This policy is a solution of Equation 33 when $\beta = 0$. In the first step Equation 36, the policy is evaluated to give the expected utility Q_g^π, then the conditional distributions for the policy per goal, $p(a_t|s_t, g)$ are computed. In the second step Equation 37 the goal is marginalised out from $p(a_t|s_t, g)$, resulting in a policy $p'(a_t|s_t)$ that minimises RGI. $p'(a_t|s_t)$ is inserted back into Equation 36 in the next iteration. Convergence of the dual-minimisation for RGI has not been proved, but in practice convergence is good for much of the range of possible values of β. The policy convergence error ϵ is defined as the maximum difference between the policies obtained in the current and previous iterations (Equation 38). The convergence test is that ϵ is below the maximum acceptable error. Suitable maximum values of ϵ were determined experimentally, see Section 4.2.2 for details.

$$\epsilon = \|\boldsymbol{\pi_{ij}} - \boldsymbol{\pi}_{ij}\|_\infty = \max \sum_j \left|\boldsymbol{\pi}_{ij} - \boldsymbol{\pi}'_{ij}\right| \tag{38}$$

where $\boldsymbol{\pi}$ is the full policy matrix having elements $\boldsymbol{\pi}_{ij} = p(a_t{=}j|s_t{=}i)$ and $\boldsymbol{\pi}'_{ij} = p'(a_t{=}j|s_t{=}i)$. When the algorithm terminates because the convergence test has been satisfied, the policy per goal that minimises RGI for the required expected utility has been found.

Comparing Equation 37 with Equation 22 from relevant information (RI), the minimum information policy for RI minimises the action entropy $H(A_t)$ by marginalising out S_t, whereas for RGI, the minimum information policy minimises the action entropy *given the state* $H(A_t|S_t)$ by marginalising out G. Thus a minimum RGI policy does not incorporate the action compression across states that is a feature of relevant information (Polani, 2011) as discussed in Section 2.4.1. There is no benefit in terms of RGI of following the same action across states. Developing a goal oriented measure that does incorporate action optimisation across states (and therefore over time) is beyond the scope of this study and left to a future investigation. However, following the example in Polani (2011) if the actions are randomised at each state, although the RGI does not change, then if $H(A_t)$ is computed, it is significantly higher than for a world with actions that are coherent across states. Even if RGI does not demand action coherence, it would perhaps be beneficial in terms of storage (and/or recall) of the policy. For these reasons, and because humans favour coherent actions, the model presented here uses the same action set in each state.

In contrast to relevant information, the minimum RGI policy optimises the policy across goals, rather than states. The dual-minimisation between Equations 36 and 37 ensures that the policy

per state for each goal is as similar as possible to the policy average across all goals. In general the agent prefers to take the same route for as many goals as possible, minimising the information bandwidth.

4.2.2 RGI policy convergence

The convergence of the dual-minimisation algorithm used to find the minimum relevant goal information policy has not been proved. However Polani et al. (2006) and Van Dijk (2013) report good convergence in practice using value iteration to evaluate the policy at each step in the dual minimisation. Finding policies for the Soho street network, and six room grid worlds were computationally heavy tasks. The number of iterations required for convergence, for a given world, increased with decreasing β. The time for each iteration, and the number of iterations, increased with the size of the state space.

Using value iteration, the first iteration of the algorithm, where value iteration must evaluate a policy that is uniform across actions at each state, took up to 10 times longer than subsequent iterations. The duration was 50 minutes on a standard PC for a typical graph ($r = 500\text{m}, |\mathcal{S}| = 418$) to converge on a policy with error $\epsilon < 0.01$ at $\beta = 3$. Figure 14 shows the progress of policy computation for the Soho street network radius $= 500m, \beta = 3$.

However when using the linear algebra approach to policy evaluation (see Section 3.4.1), the same number of iterations were required, but convergence was faster in terms of duration (Figure 14) at 23 minutes. A thorough investigation into the reasons for this performance difference is beyond the scope of this work. Python was chosen for convenience, rather than performance, using the high precision library GMPY [7] with NumPy [8] for matrix operations and a custom implementation of the value iteration algorithm. Under this regime, for the six room grid world and street map graphs used in these experiments, inverting the matrix is faster than value iteration by nearly a factor of 2, while delivering equivalent policies (i.e. within acceptable error).

When computing for the six room grid world ($|\mathcal{S}| = 627$) with $\beta \in \{4, 3, 2\}$, and for the Soho street network with larger radii, the algorithm failed to converge on a policy with $\epsilon < 0.01$, instead stabilising around a slightly higher error. This phenomenon was observed independently using either value iteration approximation or linear algebra to evaluate the policy at each iteration. Investigating the root cause of the this limit on the ability of the algorithm to converge further is beyond the scope of this work. To account for this and to ensure the policy with the lowest possible error, a second convergence test was included which terminated the algorithm once the coefficient of variation of ϵ ($c_v = \frac{\sigma}{\mu}$) had stayed below 0.05 over a window of 50 iterations, resulting in policies for the six room grid world where $\epsilon \approx 0.02$.

4.2.3 RGI per state

To compare RGI with graph centrality measures, and ultimately with hippocampal brain activation at a given time during a navigation experiment, knowing the average RGI given by Equation 39 for all states is not suitable, RGI *per state* is required.

$$RGI(s_t) \coloneqq I(G;A_t|s_t) = H(A_t|s_t) - H(A_t|G, s_t) \tag{39}$$

[7] https://pypi.org/project/gmpy2/
[8] https://pypi.org/project/numpy/

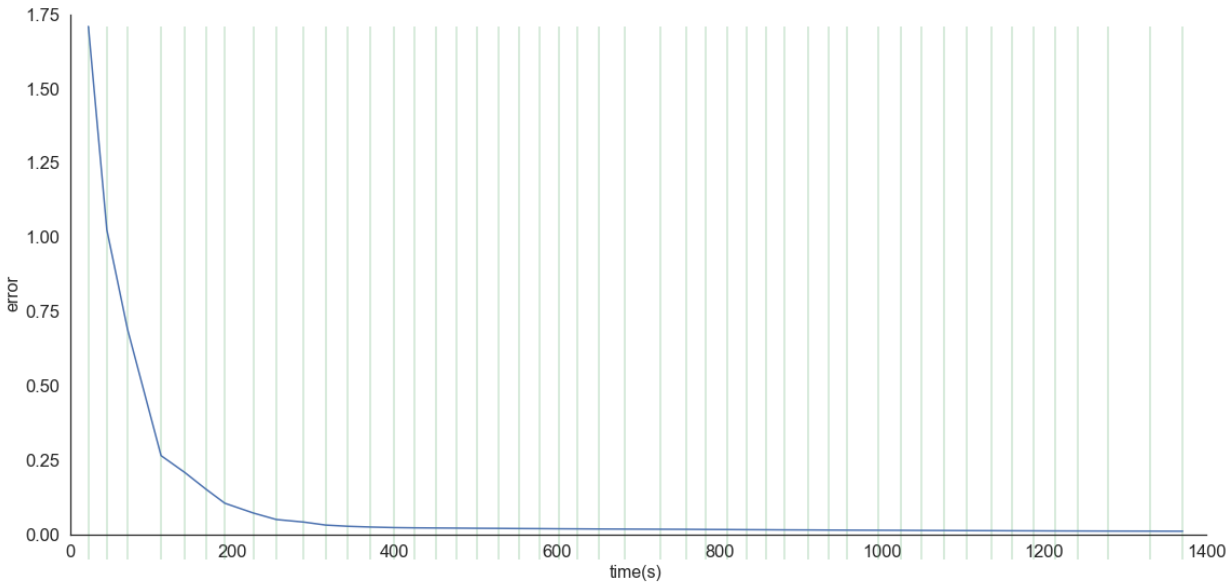


Figure 14: Policy error ϵ during policy computation, with policy evaluation by fundamental matrix, for the Soho street network radius $= 500m, \beta = 3$. Vertical lines mark the start of each iteration. The algorithm converged on a policy where $\epsilon < 0.01$ in 48 iterations with duration ≈ 1400 seconds on a standard PC.

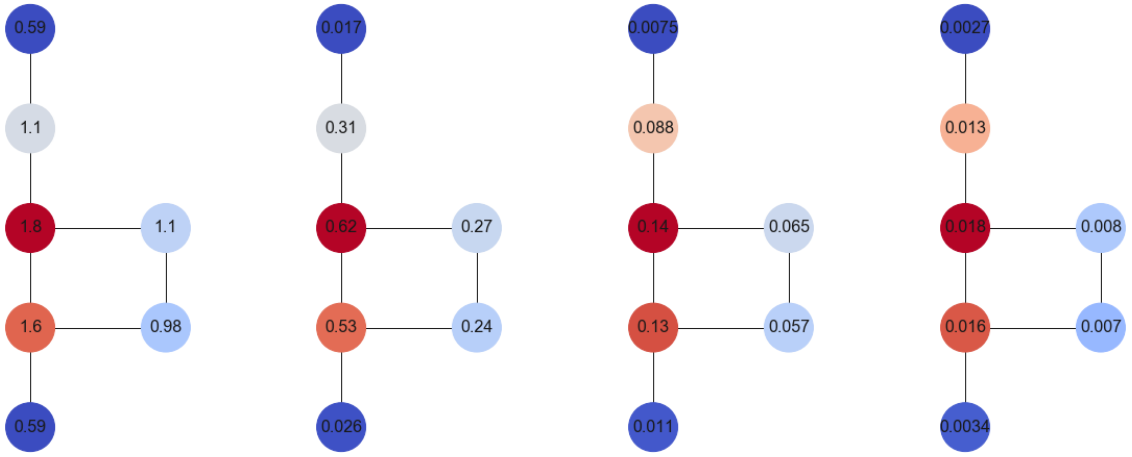
Expanding the conditional entropy terms gives:

$$I(G; A_t|s_t) = -\sum_{a_t \in \mathcal{A}_s} p(a_t|s_t) \log p(a_t|s_t) - \left[-\sum_{g \in \mathcal{G}} p(g) \sum_{a_t \in \mathcal{A}_s} p(a_t|s_t, g) \log p(a_t|s_t, g) \right] \tag{40}$$

The average relevant goal information (across all goals) for each state $RGI(s_t \in S_t)$ is computed by plugging the conditional distributions $p(a_t|s_t)$ and $p(a_t|s_t, g)$ found by the dual-minimisation algorithm into Equation 40.

Figure 15 shows the changing values of RGI as β decreases ($\beta \in \{10, 1, 0.25, 0.05\}$) for a small grid world graph. The mean RGI decreases with decreasing β, (see Figure 16 for the mean RGI/mean utility tradeoff at different values of β) but the pattern of RGI across the states remains similar. In this small world the agent can trade off a lot of information for a small loss in utility as shown by the tradeoff curve in Figure 16.

4.2.4 RGI for the small graph world



(a) $\beta = 10$ (b) $\beta = 1$ (c) $\beta = 0.25$ (d) $\beta = 0.05$

Figure 15: RGI in *bits* for the small graph world with $\beta \in \{10, 1, 0.25, 0.05\}$. The vertex colours are specific to each graph showing how the distribution of RGI across the states (reflecting the topology) remains similar despite a large change in mean RGI.

Figure 15(a) shows RGI for a maximum utility policy. Recalling that RGI is the minimum amount of information needed on average to make a decision, the state at the bottom of the graph has a low value because for almost all goals (the only counter example being where the agent is already in the goal) the choice of action is the same – to move North. At the central vertex, with RGI=1.8 *bits*, the possible goals are fairly evenly distributed amongst the actions. To act optimally, the agent needs close to the maximum 2 bits of information to choose an action. In contrast, at the state with RGI=1.6 *bits*, despite having the same set of actions to choose from, the agent can reduce the overall information required by having a policy that favours North over East and East over South, because there are more possible goals in those directions. RGI values in this small world are heavily affected by the high probability that the agent is in the goal, because there are only 7 states. As the state space increases in size this effect reduces such that "dead-end" states with only one neighbour have a low RGI, tending to zero as the number of states increases.

It is possible to manipulate the transition function during the computation of RGI to ignore the effect of the goal, resulting in a RGI of zero for dead-end states, and indeed this may be more "realistic" given that the choice of action is obvious regardless of the goal. However, by so doing the tradeoff curves shown in the next section are also affected, and since the effect of the goal on the RGI reduces with larger state spaces, leaving the goal in place for the computation of RGI (and RGIU in Section 4.3), results in a somewhat simpler formalism.

4.2.5 Information utility tradeoff

The effect of β in Equation 36 is to constrain the utility during the policy computation, permitting computation of a utility-optimal policy if required, by setting β to a high value, or computation of a policy that requires less relevant goal information, on average, by using a lower value of β. The value of β to achieve a certain ratio of utility loss to information reduction depends on the size and structure of the environment model. β is not an informational efficiency coefficient comparable

across agents and environments. However animals have evolved to optimise various biological and physical tradeoffs, and these tradeoffs do lead to sub-optimal behaviour in some circumstances, and this is modelled here by β, whose appropriate values must be discovered experimentally for each scenario.

Figure 16 shows the mean utility and mean RGI for policies computed for the small graph example used in the previous section. This curve is the characteristic tradeoff of utility for information as described by Van Dijk et al. (2010), showing the minimum information case where $\beta = 0$ and the policy is a random walk and the optimal policy where $\beta \geq 10$ above which increasing β makes no difference.

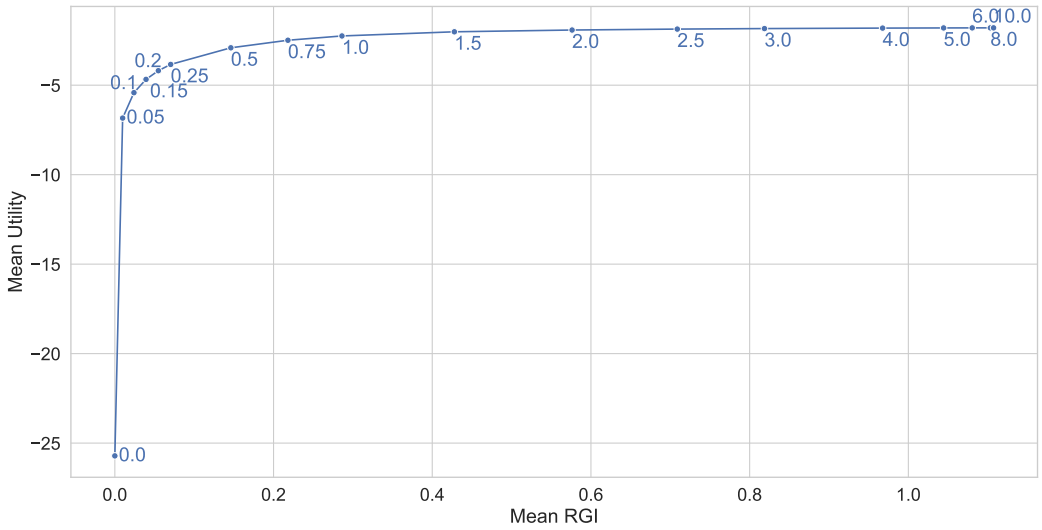


Figure 16: Mean RGI against mean utility for when computing policies for the small graph example. At $\beta \geq 10$ the policy is optimal in terms of utility. At $\beta = 0$ the policy is unaffected by the utility so the RGI is minimised regardless of cost, being a random walk. The optimal value of β depends on the relative cost of the loss in utility (extra distance walked) versus the extra cost of information processing. The information-parsimony theory indicates that organisms are likely to accept some loss of utility to reduce information processing demands, operating somewhere in the top-left region of this curve. Each combination of environment, organism and task would yield a different curve, and a different β so we cannot consider β as a universal coefficient, however an organism taking into account the diminishing returns provided by striving for optimality could operate a general strategy of achieving near optimal utility for potentially a large average saving of information processing energy cost across the full range of tasks and environments in which it operates.

4.2.6 RGI for the six room grid world

Van Dijk and Polani (2011) demonstrated that RGI uncovers important features in the environment where informational transitions occur, such as the crossing points opposite pairs of doorways. See Figure 17 which shows RGI at $\beta = 100$ for the six room grid world, reproducing Sander Van Dijk's result. These transitions are less stark in the street network, but nonetheless, as we shall see in Section 5, states (junctions) with a high RGI also tend to have high graph centrality, and the change in graph centrality has been shown to be correlated with brain activity during navigation (Javadi et al., 2017)

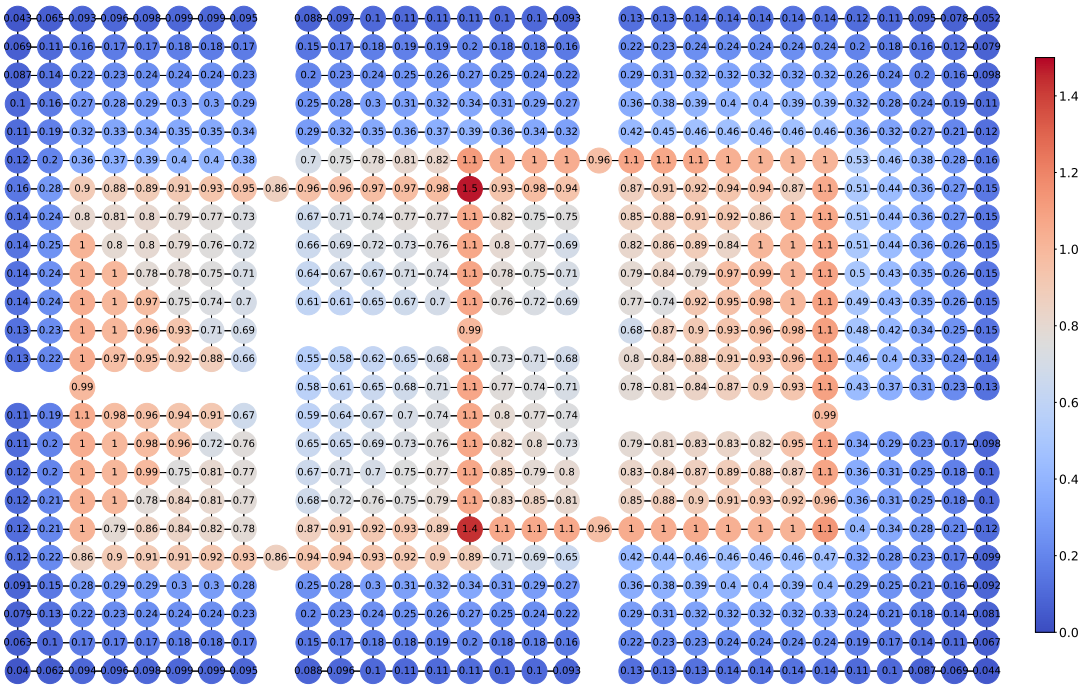


Figure 17: RGI (unweighted) for the six room grid world with $\beta = 100$ reproducing a result from (Van Dijk and Polani, 2011, p.5).

β	Mean RGI $\bar{I}$	Mean Utility $\bar{u}$	Utility Loss $\frac{\bar{u}-\max \bar{u}}{\max \bar{u}}$	RGI Reduction $\frac{\max \bar{I}-\bar{I}}{\max \bar{I}}$
100	0.519	-20.8	0	0
10	0.518	-20.8	8.95×10^{-7}	1.39×10^{-3}
4	0.498	-20.9	5.72×10^{-3}	0.0400
3	0.481	-21.3	0.0263	0.0724
2	0.445	-26.1	0.256	0.143

Table 1: Tradeoff of mean utility for a reduction in mean RGI for varying β for the six room world.

Figure 18 and Table 1 show the tradeoff of mean utility for mean RGI for policies computed for the six room grid world for $\beta \in 100, 10, 4, 3, 2$. At $\beta = 3$ a 7% reduction in RGI is traded for a 3% loss of utility, however with $\beta = 2$ a 14% reduction in RGI requires a 26% loss of utility.

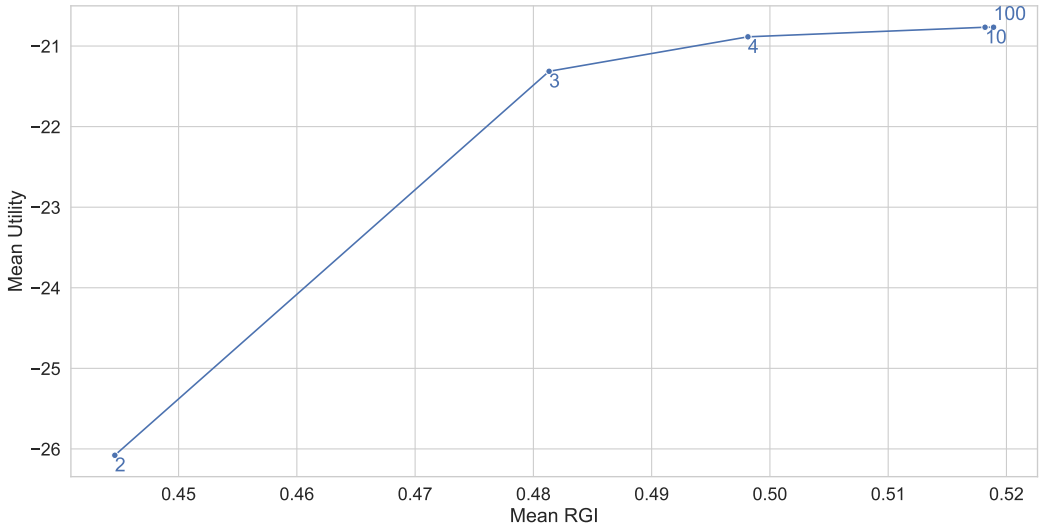


Figure 18: Mean RGI against mean utility for the six room grid world. At $\beta = 3$ a 7% reduction in RGI is traded for a 3% loss of utility, however with $\beta = 2$ a 14% reduction in RGI requires a 26% loss of utility.

4.2.7 RGI for the Soho street network

For the Soho street network at 500m radius, $\beta \geq 10$ results in an optimal policy. However, smaller values of β result in a policy that trades some utility for a reduction in the amount of information required about the goal to take decisions. Figure 19 depicts the tradeoff curve of mean utility and mean relevant goal information across all states in the 500m radius Soho street network. The computation of policies at a range of values of β for graphs of the street network at other radii show similar tradeoff curves.

Participants in the Javadi et al. (2017) study were asked questions about the optimal route, so the obvious choice would be high β when computing a policy, and subsequently RGI. However the information parsimony theory suggests that humans have evolved to minimise information processing cost where possible. For the 500m radius street network, setting $\beta = 3$ results in a policy where a 4% loss in utility is traded for a 20% reduction in RGI, which, depending on the relative energy costs of cognition and movement in the environment, may represent a useful reduction in overall energy use on average when navigating the network.

4.3 Relevant Goal Information Uptake

Relevant Goal Information $RGI(s_t)$ is the average information required about the goal at state s_t in order to take action to achieve some desired utility. An agent that forgets everything at each step would, on average, need to load this much information into working memory to make a decision. However, if information can persist in working memory, the agent would only need to load the portion of this information that was not already loaded from the previous time step. Van Dijk (2013) defines the new goal information $I_G^{new}(s_t) = I(G; A_t|E_{t-1}, s_t)$, where E is a random variable describing the full history of the agent states and actions $E_{t-1} = (S_0, A_0, S_1, A_1, ..., S_{t-1}, A_{t-1})$.

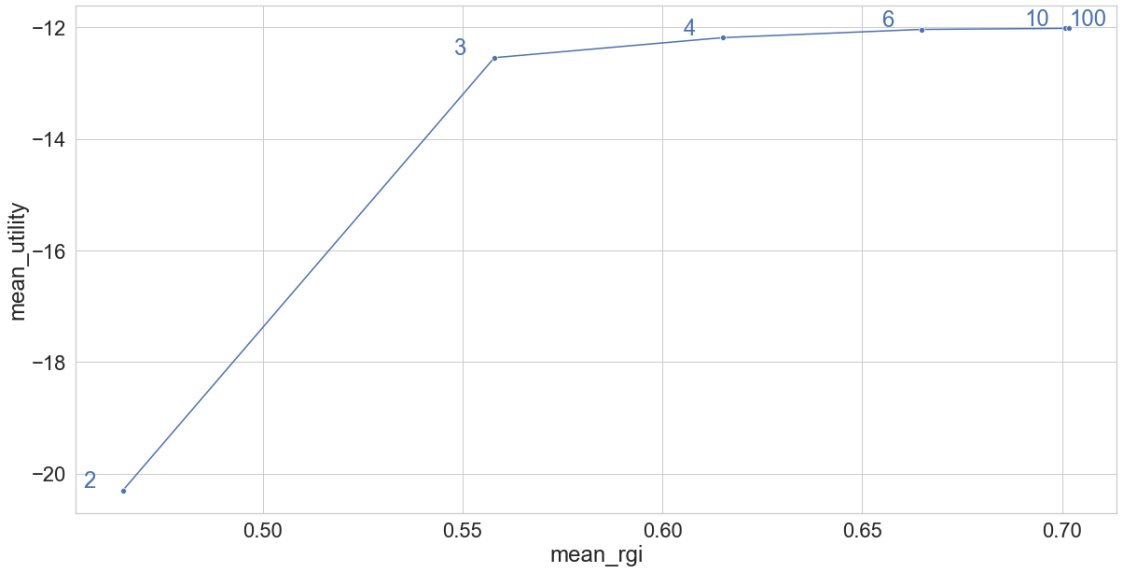


Figure 19: Mean RGI against mean utility for $\beta \in \{2, 3, 4, 6, 10, 100\}$ when computing policies for the Soho street network at 500m radius. The mean utility is optimal for $\beta \geq 10, +/- 0.001$. At $\beta = 3$ a 4% loss in utility is traded for a 20% reduction in RGI, however at $\beta = 2$ the policy results in a 41% loss of utility for 34% reduction in RGI.

In the present model the transition function is deterministic, so we can disregard actions from the history and just consider states. Nonetheless the curse of dimensionality strikes – computing the full distribution for all possible previous state trajectories quickly becomes intractable. On average, the further we peer into the past, the less goal information required at that state is relevant now, so I^{new} can be estimated by limiting the history to a finite number of steps[9]. I define the random variables $S^n_{t-1} = (S_{t-n}, ..., S_{t-1})$ to represent the state history distribution over n steps, and name the time-limited estimate of I^{new} Relevant Goal Information Uptake (RGIU).

$$RGIU^n(s_t) := I(G; A_t | S^n_{t-1}, s_t) = H(A_t | S^n_{t-1}, s_t) - H(A_t | G, S^n_{t-1}, s_t) \tag{41}$$

RGIU can also be thought of as the reduction in the action uncertainty provided by knowledge the goal given that n previous states are known. RGIU is the *new* information needed about the goal in a given state, given the trajectory (n steps), hence it is dubbed "uptake". Results from Javadi et al. (2017) suggest that humans do indeed update a navigational plan at street entries rather than planning from scratch while navigating: "Because we found that hippocampal activity reflected the change in degree centrality, not the raw degree centrality, a model in which the hippocampus only processes future paths the moment a street is entered is not consistent with our data. Rather, our data agree with a model in which the hippocampus simulates possible paths throughout the journey, with the hippocampal activity we observed reflecting the increase or decrease in the number of potential future paths to be re-activated as each new street is entered" (Javadi et al., 2017).

The formalism for RGIU presented here only includes a random variable for actions in the current state A_t. This is made possible by the deterministic transition function, otherwise the

[9]Please see Appendix A Figure 37 which shows RGIU with 3 steps of history weighted by the probability that an agent is in each state over all time $p(s)$ which shows a similar characteristic pattern to I^{new} in Van Dijk (2013)

previous action variables $A_{t-1}, .., A_{t-n}$ would need to be added to Equation 41. As discussed in Section 4.2 with regards to RGI, one upshot of not including the historic sequence of actions in the formalism is that coherent actions are also not required for a minimum RGIU policy. Once again incorporating the action variables and assessing the influence of action coherence is deferred to a future study.

4.3.1 Computing RGIU

To compute RGIU from the conditional entropies in Equation 41, we need to know the distributions of the agent state over n previous time steps modelled by $S_{t-1}^n = S_{t-n}, ..., S_{t-1},$. The Markov Chain of states for an agent seeking a goal, is defined by the state transition matrix $\mathbf{P} = p(s_{t+1}|s_t)$. To compute these distributions, if we know the state distribution vector at the earliest step $\mathbf{s_0} = S_{t-n-1}$, as we have seen in Section 2.2.1, the next n state distributions can be found by repeatedly applying the transition matrix.

$$\mathbf{s_{k+1}} = \mathbf{s_k}\mathbf{P} \tag{42}$$

How do we determine the initial state distribution vector $\mathbf{s_0}$? In other words: what is the state probability distribution S if we don't know the time t? We know from Section 2.2.1, that we can find the (time-independent) stationary distribution for a regular Markov chain, however the goal is an absorbing state – once it is in the goal it will stay there. For an agent executing any policy where it is not actively avoiding the goal, the probability of being in the goal tends to 1 over time i.e. $\lim_{t\to\infty} \Pr(S_t{=}g) = 1$, thus the Markov chains induced are not regular. We need a distribution that represents the agents' state *just during the period of interest*, as it moves towards the goal. The solution to this problem is to adapt the chains so only the period of activity before reaching the goal is considered – to ensure regularity without compromising the model – so that a useful stationary distribution can be computed to use as the initial state distribution s_0.

The same trick was employed as by Van Dijk (2013) to modify the transition function such that, in the step after arriving at the goal, the agent is "teleported" to a (uniformly distributed) random location $\forall s_{t+1} \in \mathcal{S}, \Pr(S_{t+1}{=}s_{t+1}|s_t{=}g) = \frac{1}{|\mathcal{S}|}$. With this modification in place, the induced Markov chains are regular, and the stationary distribution vector $\mathbf{w}$ (the average distribution of agent states before arriving at the goal) can be readily computed.

The joint distribution of previous states and current state $p(S_{t-1}^n, s_t) = p(S_{t-n}, ..., S_{t-1}, s_t)$, is obtained by setting $\mathbf{s_0} = \mathbf{w}$, and iterating Equation 42, setting $k = 0$ and iterating until $k = n$.

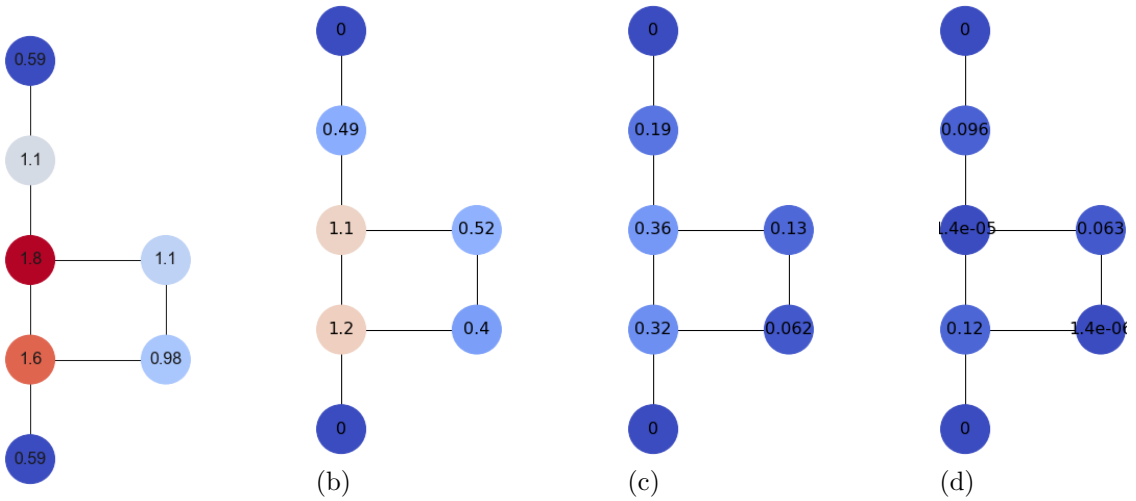
With the state history available it is straightforward (although computationally expensive for a large state space) to obtain the full joint distribution of state histories and actions at the current state over all goals, from the the policy per goal $p(a_t, |s_t, g)$, and assuming a uniform distribution of goals $p(g) = \frac{1}{|\mathcal{G}|}$.

$$p(A_t, S_{t-1}^n, G|s_t) = \prod_{g,a_t} p(S_{t-1}^n|s_t, g)\, p(a_t, |s_t, g)\, p(g) \tag{43}$$

With the full joint in hand the conditional entropies in Equation 41 can be computed to obtain $RGIU^n(s_t)$.

4.3.2 RGIU for the small graph world

Figure 20 shows RGI (a) for the small graph compared to RGIU with 1, 2 and 3 steps of history, all at $\beta = 10$.



(a) RGI (b) $RGIU_{n_{steps}=1}$ (c) $RGIU_{n_{steps}=2}$ (d) $RGIU_{n_{steps}=3}$

Figure 20: RGI and RGIU in *bits* for the small graph world with $\beta = 10$. RGIU for $n_{steps} \in \{1, 2, 3\}$ Vertex colours are consistent across all four graphs, showing the reduction in RGIU as the number of steps increases.

In general, knowing the history reduces uncertainty about the goal, and for all states in this small example world where β is unchanged, $\forall s \in \mathcal{S}, RGI(s) > RGIU_{n_{steps}=1}(s) > RGIU_{n_{steps}=2}(s) > RGIU_{n_{steps}=3}(s)$. Unless the agent has a random policy, S_t and S_{t-1} are not independent – knowing S_{t-1} reduces uncertainty about S_t hence for most interesting worlds (more than 5 states, reasonably connected):

$$I(G;A_t|S_t) \geq I(G;A_t|S_{t-1}, S_t) \tag{44}$$

and

$$I(G;A_t|S_{t-1}, S_t) \geq I(G;A_t|S_{t-2}, S_{t-1}, S_t) \tag{45}$$

With more steps of history the mutual information reduces further. For some states in the limit of t the RGIU is zero, but for others some information is needed to make a decision. A long memory could confer advantages for an agent, but must also come with a cost, suggesting a new tradeoff of memory and processing cost. Further investigation, and rigorous definition, of this phenomenon is beyond the scope of this study and is left to a future work.

For "dead-end" states which have only one neighbouring state (top and bottom states in the small graph), the choice of action is obvious. If the agent has a history at all, it must have been in this same state in the previous time step, in which case this is the goal – or in the previous time step it was in the neighbouring state, in which case this is the goal. No extra information besides the knowledge of a single step of history is needed to make a decision. This is the goal and the agent will wait here, consequently the RGIU for dead-ends is precisely zero.

For states with two neighbours, The choice of action is usually obvious from the history - an agent following an optimal policy will not backtrack, so will prefer the action that takes it away from the previous state. However, once again, the small size of this example graph means that there is a relatively high probability that the current state is the goal. For this reason "corridor" states (with two neighbours) do require the agent to take in some new information to decide whether to wait (because this is the goal) or to move on. As we shall see from the street network graph, as the size of the state space increases, the RGIU for states with two actions tends to zero, because the state in question, at any particular time, is decreasingly likely to be the goal.

4.3.3 RGIU for the six room grid world

Figure 21 shows RGIU with $\beta = 100$ for 1 and 2 steps of history. With this high value for β the policy is optimal in terms of utility. The mean RGIU at 2 steps is lower, but as we can see from the visualisation, the distribution of RGIU across the state space has similar characteristics.

For states towards the edges of the rooms, the average (for all possible goals) new information required to choose an optimal route is low – for most goals the same action can be taken. In the states that represent corridors between the rooms, the RGIU is always zero because the history perfectly predicts the action, if the agent has entered the corridor, the goal must be there or beyond, an agent with an optimal utility policy would never turn back.

There are six states with clearly higher RGIU, which lie on the crossing-points of lines from the doorways. At these states, because most goals are in a different room, the agent must usually head for the correct doorways. Knowing the history does not predict the goal so well in these states so on average the RGI uptake is relatively high.

Recalling that Van Dijk defined I_G^{new} as the new relevant goal information with full knowledge of the history, $RGIU_{n_{steps}=\infty} = I_G^{new}$. With the full history, and weighting I_G^{new} by $p(s_t)$, the states inline with and close to the doorways show the information transfer taking place at the doorways (Van Dijk and Polani, 2013, p.13). For comparison, Figure 37 in Appendix A shows $RGIU_{n_{steps}=3}$ at $\beta = 100$ (optimal policy in terms of utility). Although RGIU does rise in states close to the doorways, the effect is not as pronounced as demonstrated by Van Dijk, and the crossing points in the rooms opposite doorways are much more pronounced. The reasons for this are beyond the scope of this work to analyse, but could be due to the limited number of steps.

Please see Section 5 for the correlation of RGIU in the six room world with the graph centrality measures.

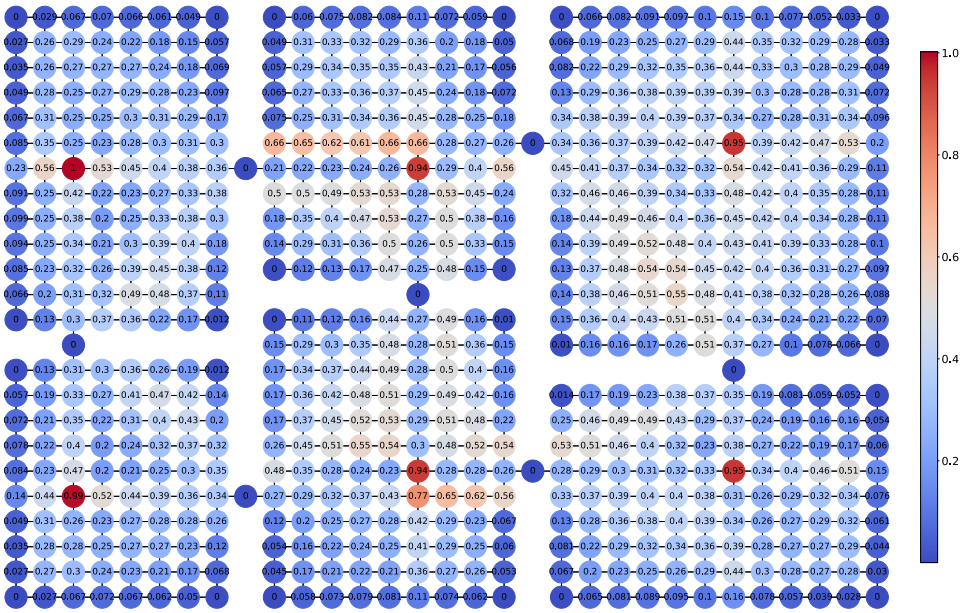
4.3.4 RGIU for the Soho street network

Figure 22 shows RGIU for the Soho street network for varying β and number of steps. RGIU is the *new* relevant goal information required on average for all goals. In general the RGIU reduces with increasing number of steps, but the reduction is not the same for all states. For junctions where the topology of the network is such that in general the additional history of the agent better predicts the location of the goal, the RGIU will drop relative to states where adding more steps of history does not reduce the uncertainty about the goal so much. Junctions with high RGIU at 3 steps of history will require more information to be loaded into working memory which may be detectable in brain activation of a different region to states where the RGI or empowerment is high. For a comparison of RGIU, RGI and centrality please see Figure 32.

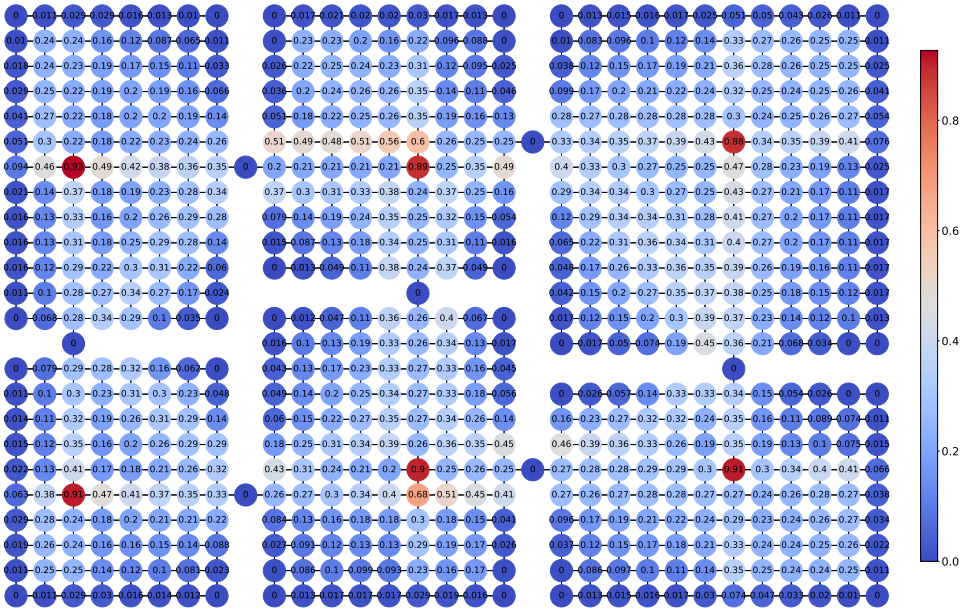
Figure 23 shows a small central section of the 500m radius Soho street network with RGI and 1-step RGIU computed at $beta = 10$. The comparison demonstrates how RGIU reflects different

aspects of the topology of the network. Corridor states with two neighbours have a very low RGIU, compared with RGI for the same state (e.g. state U319). Certain states with high RGI (e.g N134, N143) have relatively low RGIU. At these states, although the average total goal information required at that state is high, the average amount of *new* goal information is not so high compared with neighbouring states. If the cost of loading and unloading goal information into working memory is indeed significant, it is possible that, when measured navigating these junctions, activation in some region of the brain will scale with RGIU.

The radius of the map affects the values of Relevant Goal Information for vertices (states/junctions) of interest, however the effect is not monotonic. For some junctions, the RGI drops with increasing radius considered, for others RGI increases with increasing radius, reflecting the shifting topology of the graph as more nodes are added to the periphery. RGIU exhibits a similar changing pattern of values as the radius of the map changed.

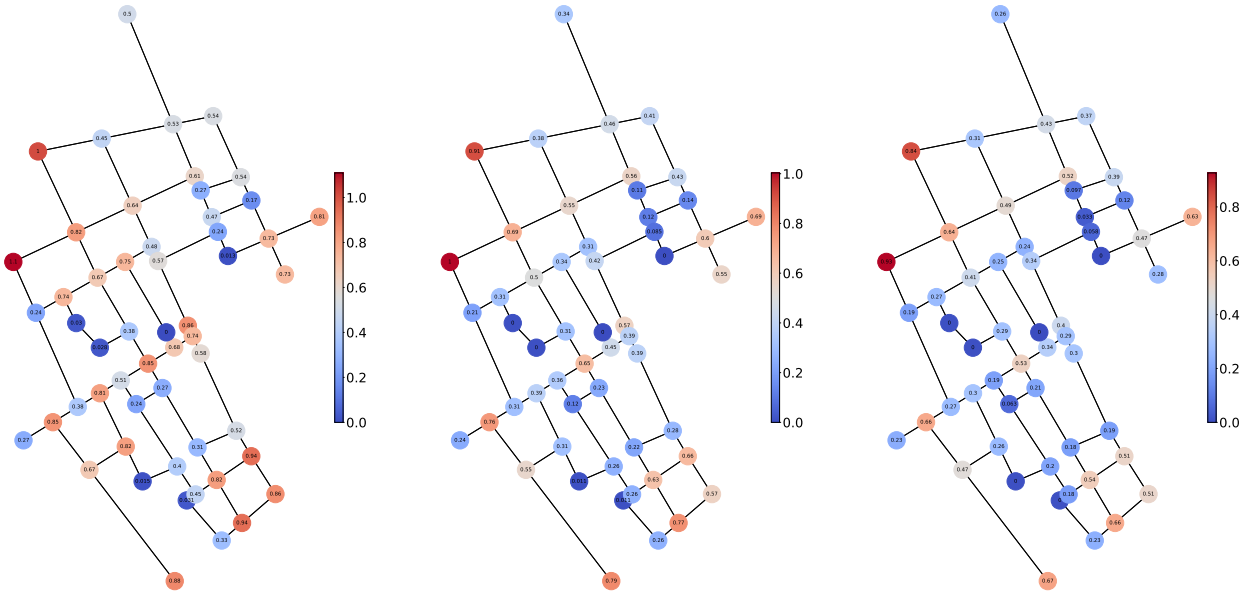


(a) RGIU $\beta = 100, n_{steps} = 1$

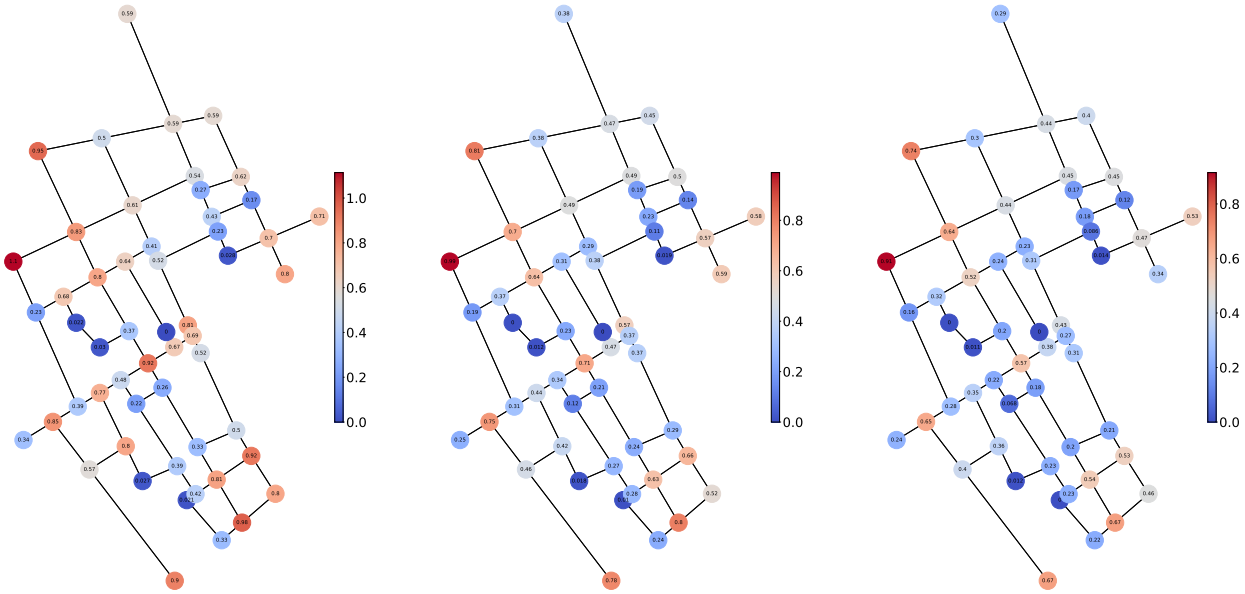


(b) RGIU $\beta = 100, n_{steps} = 2$

Figure 21: RGIU $\beta = 100, n_{steps} \in \{1, 2\}$ for the six room grid world. Colour ranges are specific to each graph. The distribution of RGIU with 2 steps of history is similar to the distribution with a single step, although values are lower, as expected given the extra information about the goal provided by the extra knowledge of the prior trajectory. The states with high RGIU indicate where an agent must load more information into working memory, on average across all goals, because the previous state does not predict the goal very well. For architects and interior designers these points may indicate where signage could be deployed to alleviate the cognitive burden of wayfinding in buildings.

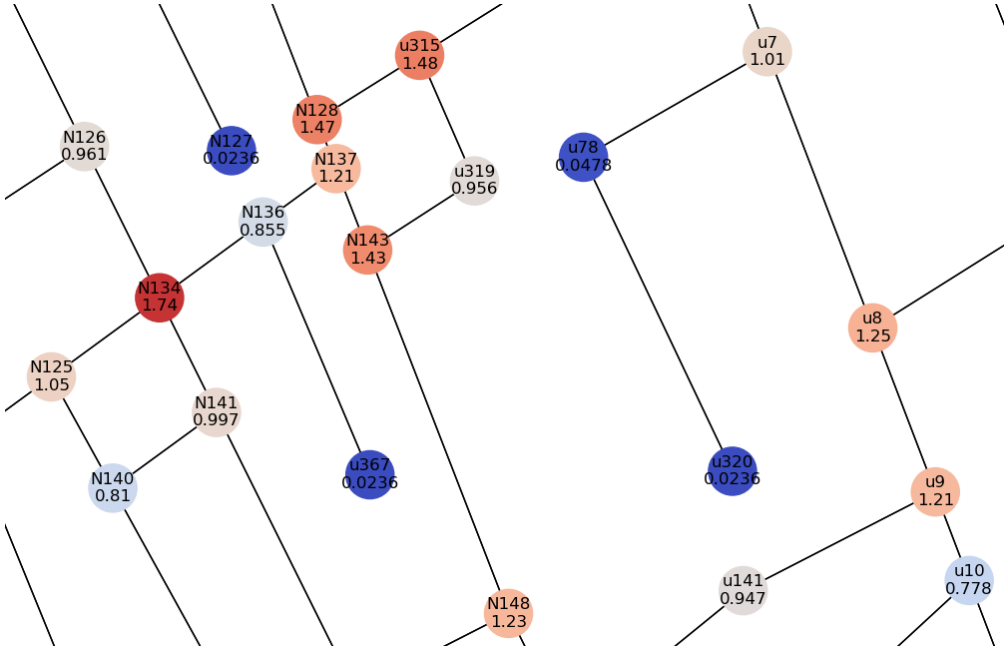


(a) RGIU $\beta = 100, n_{steps} = 1$ (b) RGIU $\beta = 100, n_{steps} = 2$ (c) RGIU $\beta = 100, n_{steps} = 3$

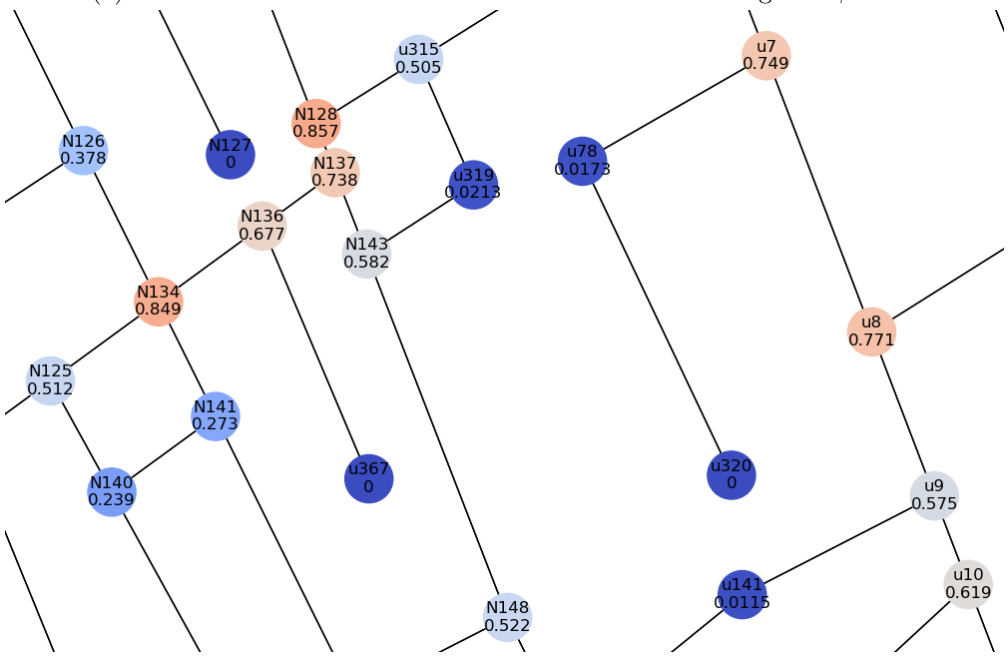


(d) RGIU $\beta = 3, n_{steps} = 1$ (e) RGIU $\beta = 3, n_{steps} = 2$ (f) RGIU $\beta = 3, n_{steps} = 3$

Figure 22: RGIU $\beta \in \{100, 3\} n_{steps} \in \{1, 2, 3\}$ for the Soho street network at 500m radius showing only vertices of interest ($n = 31$). The pattern of high and low RGIU is similar for varying β and n_{steps} but the relative level of some junctions does change. The two junctions top-left stay relatively high with increasing steps, while other junctions, such as the central crossroads, reduce in relative RGIU. Colour ranges are specific to each graph. Vertices not included in the Spiers lab study are not shown, so vertices that are shown maybe more connected than they appear here.



(a) Section of Soho street network at radius 500m showing RGI $\beta = 10$



(b) Section of Soho street network at radius 500m showing RGIU $\beta = 10, n_{steps} = 1$

Figure 23: Comparison of RGI and RGIU for the Soho street network showing how RGIU illuminates different features of the topology of the network. Corridor states (states connected to two other states) have a low RGIU - knowledge of the previous state very accurately predicts the goal. This makes intuitive sense - someone having chosen a goal a few streets away does not need to think about which way to go when turning a corner.

4.4 Additional steps of history reduce uncertainty less for informationally parsimonious policies

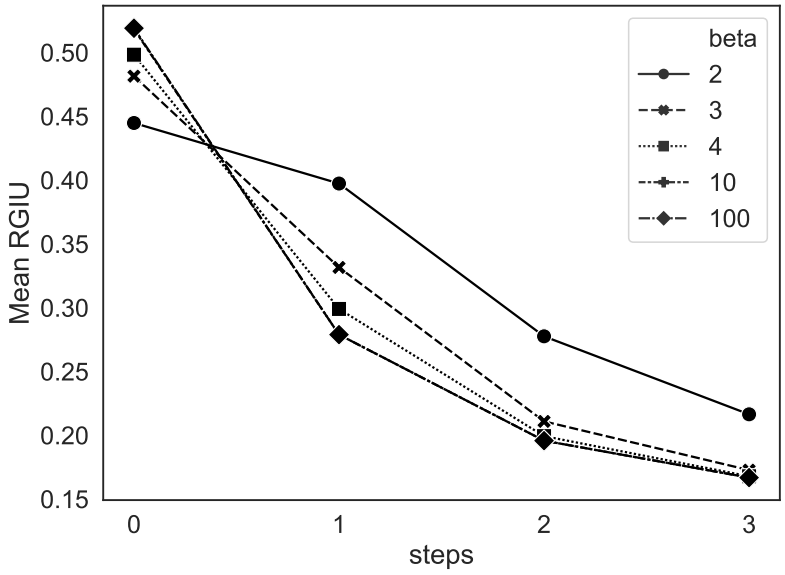


Figure 24: Mean RGIU at $\beta \in 100, 10, 4, 3, 2$ and $n_{steps} \in \{0, 1, 2, 3\}$ for the six room grid world. Note that $RGIU_{n_{steps}=0} = RGI$.

Figure 24 shows the mean RGI ($RGIU_{n_{steps}=0}$) and mean RGIU for 1, 2 and 3 steps of history computed for the six room grid world. For a specific policy, increasing the number of steps of history, reduces uncertainty about the goal. As β decreases the mean RGI also decreases (this is the tradeoff described above), but the reduction in RGI provided by each step of history is less. Although the mean RGI is lower for $\beta = 2$ than for $\beta = 100$, the policy per goal $p(a_t|s_t, g)$ has a higher average entropy – there is more uncertainty in the actions, and this means that knowing the history provides a lower reduction in uncertainty about the goal. Exploring the implications of this is beyond the scope of this study and is reserved for future work.

5 Results

5.1 Summary of results

The aim of this study is to compare the information-theoretic measures with the graph-theoretic measures in graph representations of interior space and the Soho street network. The results are consistent with the hypothesis in Section 1.7 showing a strong correlation between the information-theoretic measures and the graph-theoretic quantities measured on graph representations of interior space and the street network of Soho.

For the six room grid world, empowerment is very strongly correlated with degree centrality, closeness centrality and betweenness centrality, relevant goal information is strongly correlated with closeness centrality and betweenness centrality, and relevant goal information uptake is strongly correlated with degree centrality. These results suggest that the information theoretic measures could be used for agent-based models of navigation through indoor spaces.

For the Soho street network, measuring centrality with the primal graph, empowerment is strongly correlated with degree centrality, closeness centrality and betweenness centrality. When the centrality measures are computed using Space Syntax and the dual graph, empowerment is also very strongly correlated with degree centrality and strongly correlated with closeness centrality, but not correlated with betweenness centrality. Relevant goal information and relevant goal information uptake are strongly correlated with degree centrality, somewhat correlated with closeness centrality and not correlated with betweenness centrality. The centrality measures are commonly used by city planners to predict macroscopic movement patterns, so these results suggest that the information-theoretic measures would also be useful in this context. Furthermore the correlation with the Space Syntax measures, especially the correlation of change in empowerment with the change in degree centrality, indicates that these measures may predict brain activation and this hypothesis will be tested with a direct comparison with the fMRI data from the Soho navigation study in further work that is in progress at the time of writing.

5.2 Parameters of measures computed and tested for correlation

All measures were computed for each radius $r \in \{300, 400, 500, 600, 700\}$. Empowerment was computed for 1 to 9 steps $n_{steps} \in [1, 9]$. Relevant goal information (RGI) was computed for the following range of utility/information trade-off coefficients $\beta \in \{100, 10, 8, 6, 4, 3, 2\}$ Relevant Goal Information Uptake (RGIU) was computed for all combinations of the RGI parameters r, β and for 1, 2 or 3 steps of history $n_{steps} \in \{1, 2, 3\}$. All correlation coefficients reported in this section are Pearson coefficients.

5.3 Correlation of Empowerment with primal graph centrality

5.3.1 Six Room Grid World

Figure 25 shows the correlation between empowerment at 1 to 9 steps and the three centrality measures for the six room grid world. Unsurprisingly empowerment is tightly correlated with degree centrality ($\mathfrak{E}_{n_{steps}=1} = \log_2 C_D$), with the amount of correlation decreasing as the number of steps in the look-ahead horizon of empowerment increases. Correlation with closeness centrality increases with more steps, because empowerment at a small number of steps is a local quantity but as the horizon expands towards the size of the graph, empowerment becomes a global quantity. Empowerment is increasingly correlated with betweenness centrality as the number of steps

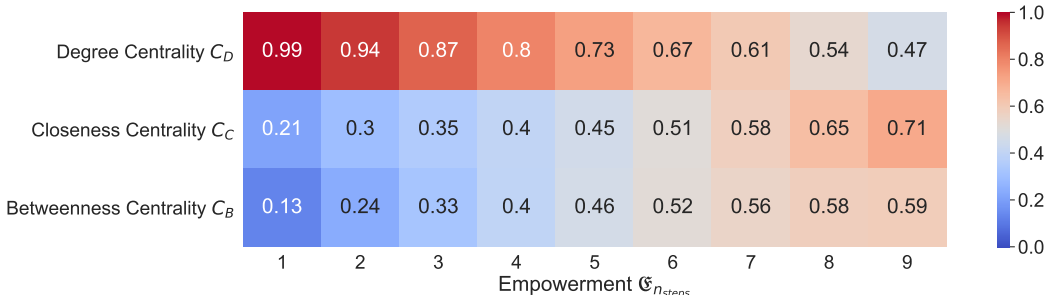
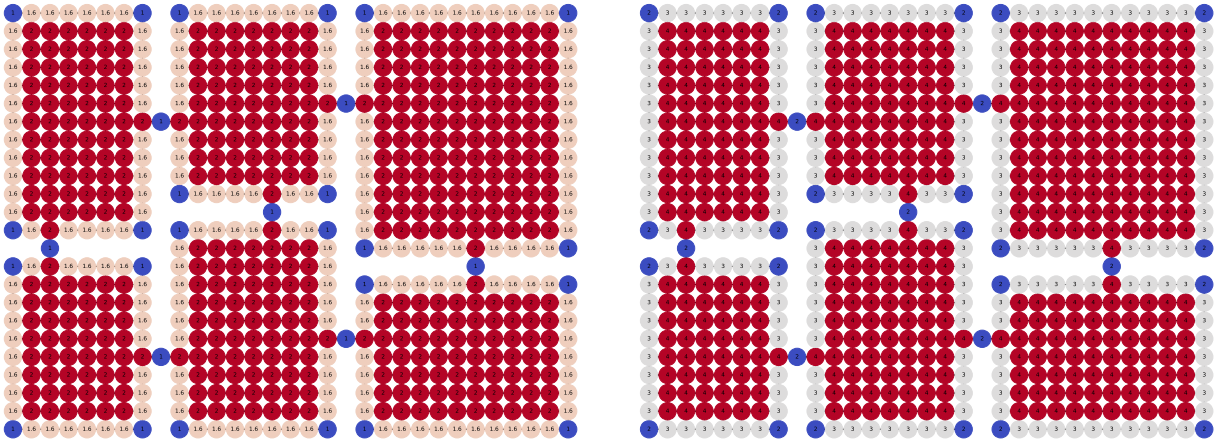


Figure 25: Correlation of Empowerment $\mathfrak{E}_{n_{steps}\in[1,9]}$ with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for a six room grid world.

increases, however the rate of improvement of the correlation slows. This effect is related to the size of the rooms in the six-room world. Betweenness centrality is highest for states that the agent is most likely to traverse for all possible journeys, so the "doorway" between the rooms are clearly identified. As the number of steps increases, there are more states where the empowerment horizon makes the of states beyond the doorway accessible to the computation.

Figure 26 shows empowerment at 1 step, 5 steps and 9 steps for a six room grid world, alongside degree centrality, closeness centrality and betweenness centrality for the same world. Degree centrality and 1-step empowerment are essentially the same, with the colour range difference in Figures 26a and 30b due to 1-step empowerment being the log of degree.



(a) 1 step empowerment $\mathfrak{E}_{n_{steps}=1}$

(b) degree centrality C_D

(c) 5 step empowerment $\mathfrak{E}_{n_{steps}=5}$

(d) closeness centrality C_C

(e) 9 step empowerment $\mathfrak{E}_{n_{steps}=9}$

(f) betweenness centrality C_B

Figure 26: Comparison of empowerment $\mathfrak{E}_{n_{steps}\in\{1,5,9\}}$ with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for a six room grid world. Colour ranges are specific to each graph. 9-step empowerment and betweenness centrality are correlated ($r = 0.59$) but the visualisation reveals that the distribution of values looks markedly different, which may prove important when predicting the behaviour of people in interior spaces.

5.3.2 Soho Street Network

Figure 27 shows that empowerment for the vertices of interest ($n = 31$) in the Soho street network at 500m radius shows strong correlation with degree centrality, with the amount of correlation decreasing as the number of steps decreases (Pearson r=0.97 at 1 step to r=0.72 at 9 steps). Empowerment is also strongly correlated with closeness centrality, with two minima (Pearson r=0.79 at 1 step and 7 steps) and two maxima (Pearson r=0.85 at 4 steps and r=0.83 at 9 steps). Correlation with betweenness centrality has a maximum of $r = 0.82$ at 5 and 6 steps, which drops to 0.7 at 9 steps and 0.54 at 1 step. Figure 28 shows the visualisations of empowerment at 1, 3 and 5 steps and the three centrality measures for the vertices of interest in the graph of the Soho street network at 500m radius.

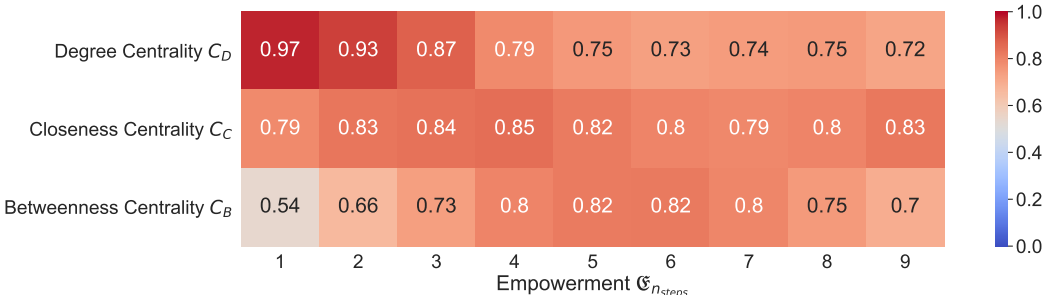
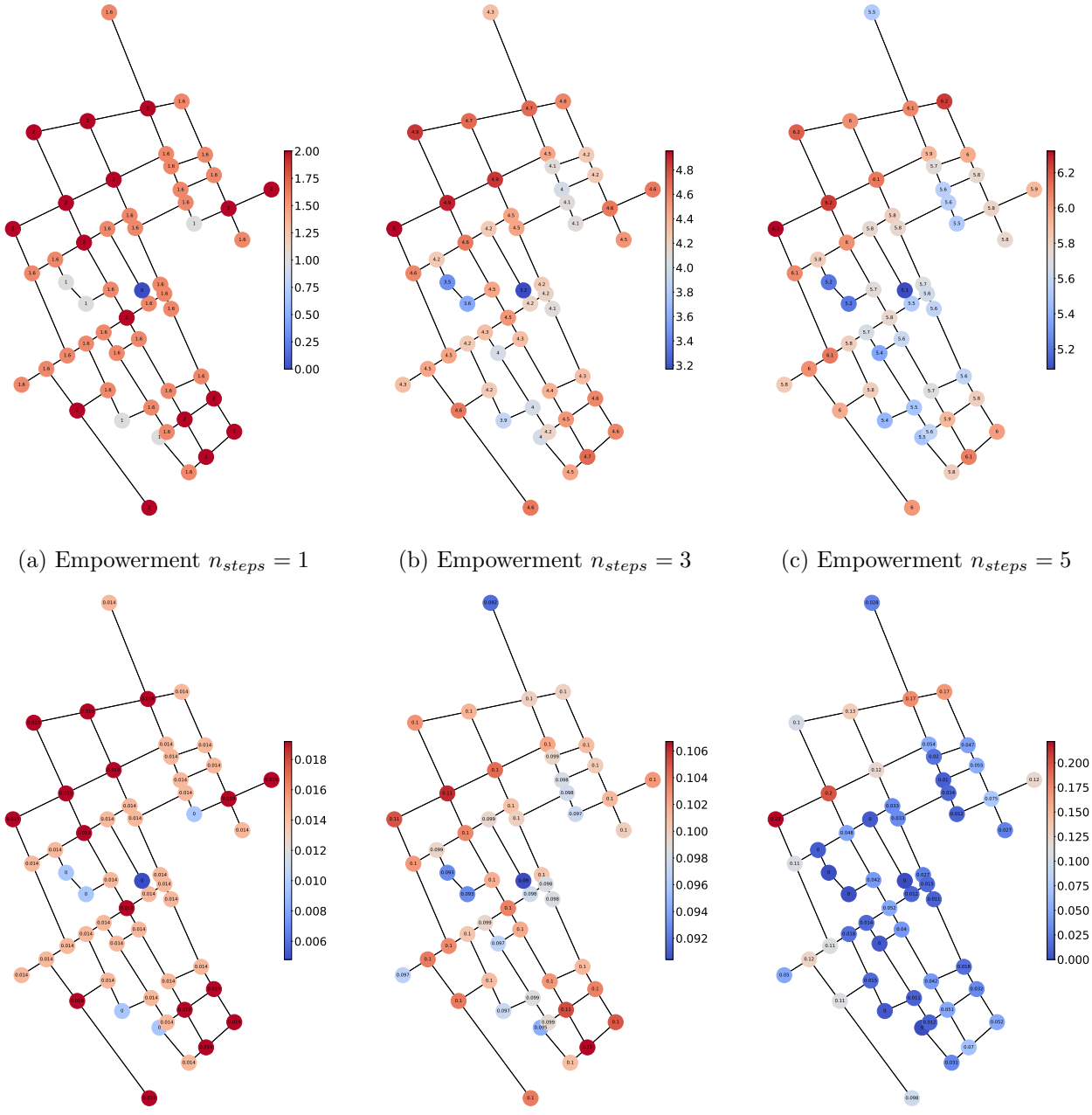


Figure 27: Correlation of Empowerment $\mathfrak{E}_{n_{steps} \in [1,9]}$ with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the vertices of interest ($n = 31$) in the Soho street network at 500m radius.



(a) Empowerment $n_{steps} = 1$ (b) Empowerment $n_{steps} = 3$ (c) Empowerment $n_{steps} = 5$

(d) Degree Centrality (e) Closeness Centrality (f) Betweenness Centrality

Figure 28: Empowerment $n_{steps} \in \{1, 3, 5\}$ for the Soho street network at 500m radius, compared with Degree Centrality, Closeness Centrality and Betweenness Centrality. Only junctions used in Javadi et al. (2017) are shown. Colour ranges are specific to each graph. Visually the distribution of empowerment at 5 steps resembles the distribution of closeness centrality, however it is equally tightly correlated with betweenness centrality.

5.4 Correlation of RGI and RGIU with primal graph centrality

5.4.1 Six Room Grid World

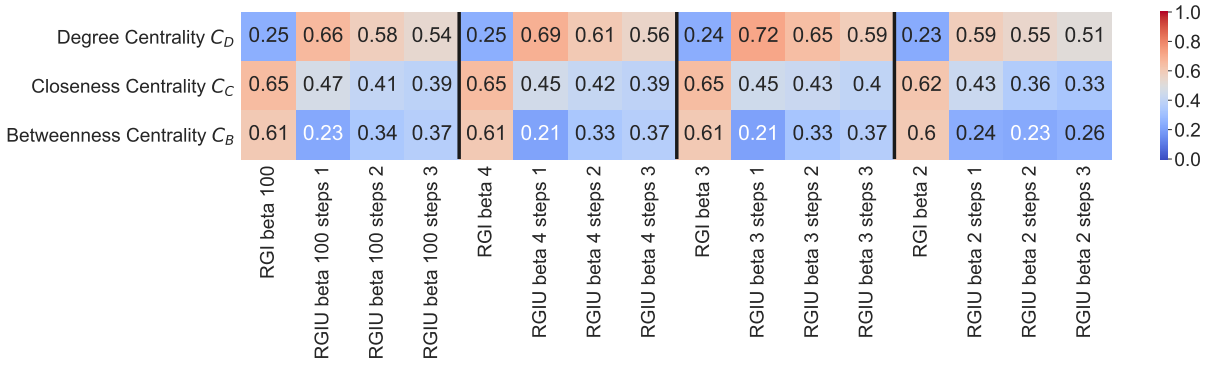
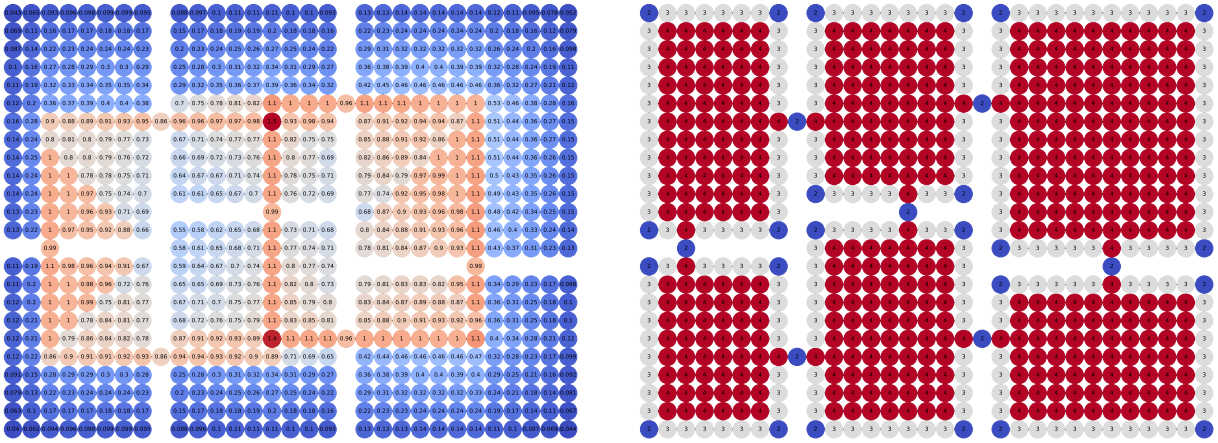


Figure 29: Correlation at $\beta \in \{100, 10, 4, 3, 2\}$, of RGI and RGIU $n_{steps} \in [1, 3]$ with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the six room grid world.

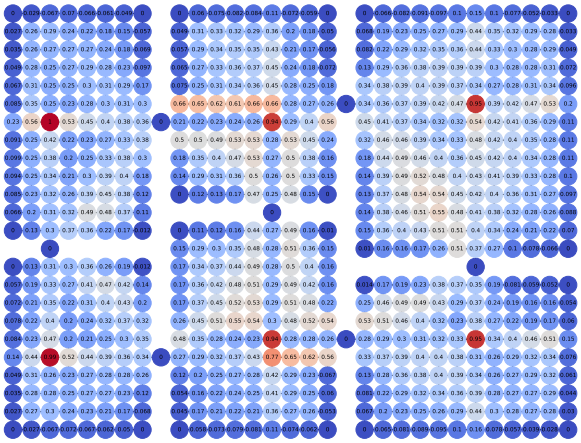
The correlation between RGI and RGIU and graph centrality, for the six room world, is less than for empowerment. RGIU with 1 step of history is somewhat correlated with Degree Centrality ($r = 0.72$ when $\beta = 100$), and RGI is somewhat correlated with Closeness Centrality ($r = 0.65$ when $\beta = 100$) and Betweenness Centrality ($r = 0.61$ when $\beta = 100$).

Figure 30 provides a visual comparison of RGI, RGIU and the three centrality measures. The relatively low correlation compared to empowerment is unsurprising given the lack of visual similarity between RGI, RGIU and centrality. However RGI weighted by the state probability distribution $p(s_t)$ shown in Figure 37 in Appendix A visually more closely resembles the pattern of betweenness centrality, which is unsurprising because the shortest path betweenness centrality of a state is proportional to the probability of an agent being in that state if it walks the shortest path between all pairs of vertices in the graph. The implications of using weighted RGI are beyond the scope of this study and are deferred to future work.

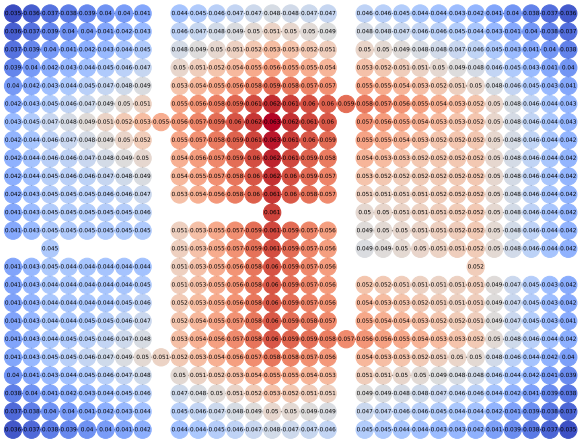


(a) RGI $\beta = 100$

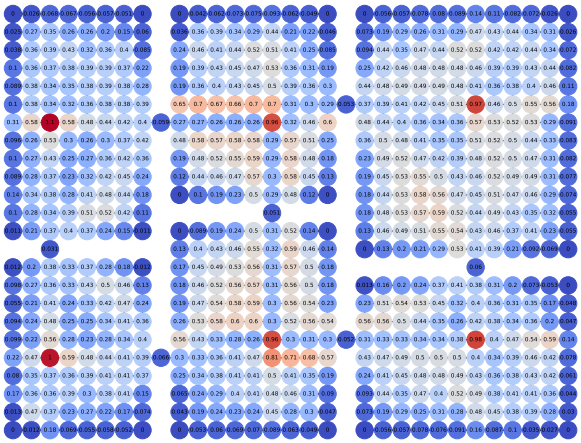
(b) degree centrality C_D



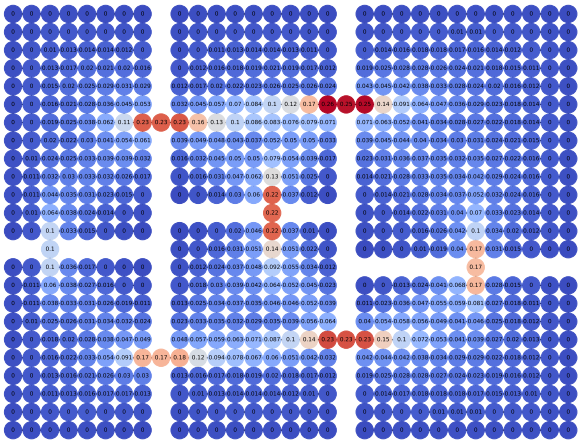
(c) RGIU $\beta = 100, n_{steps} = 1$



(d) closeness centrality C_C



(e) RGIU $\beta = 3, n_{steps} = 1$



(f) betweenness centrality C_B

Figure 30: Comparison of RGI $\beta = 100$ and RGIU $\beta \in \{100, 3\} n_{steps} = 1$ (both unweighted) with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the six room grid world. Colour ranges are specific to each graph. The comparison shows that even though RGI is correlated with betweenness centrality, and closeness centrality it has a markedly different distribution to either. The rectangular nature of the pattern of higher RGI values is an artefact of the model only allowing four movement actions.

5.4.2 Soho street network 500m radius

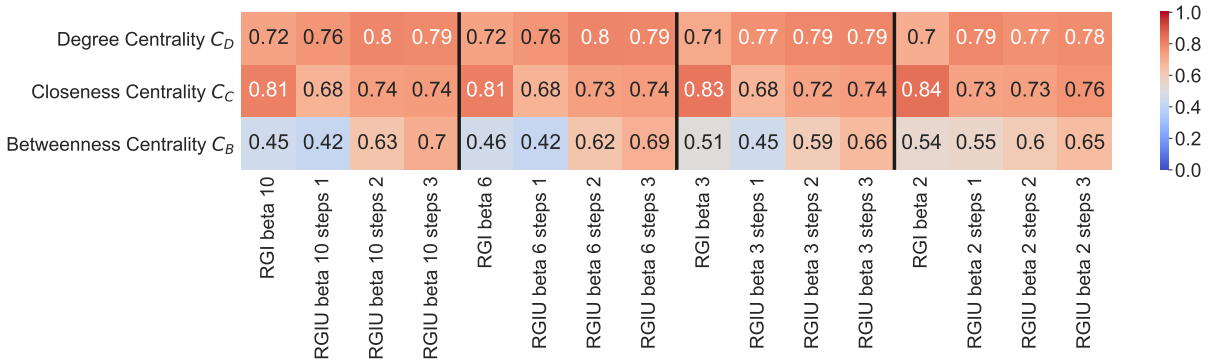
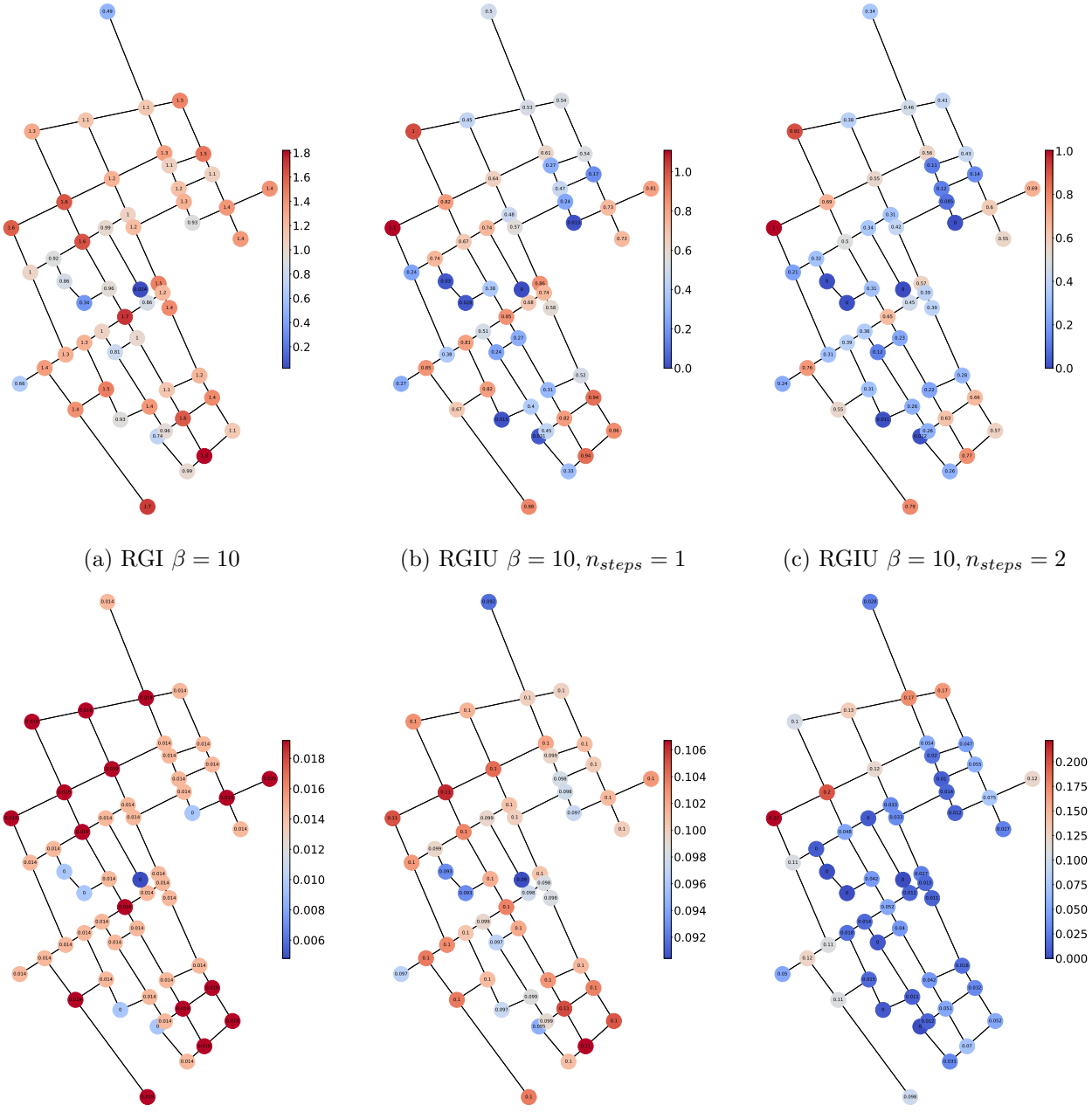


Figure 31: Correlation with $\beta \in \{10, 6, 3, 2\}$ of RGI, and RGIU $n_{steps} \in [1, 3]$ with primal graph measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the vertices of interest ($n = 31$) in the Soho street network at 500m radius. Correlation is substantially stronger across measures for the street network than for the six room grid world shown in Figure 29

Correlation for RGI and RGIU is stronger for the 31 vertices of interest in the Soho street network than for the six room grid world. RGI at high beta correlates with Degree Centrality at Pearson $r = 0.72$ and Closness Centrality at $r = 0.81$ for the Soho street network. RGIU at high beta with 3 steps of history correlates with Degree Centrality ar $r = 0.8$, Closeness Centrality at $r = 0.74$ and Betweenness Centrality at $r = 0.7$. Changing β and/or the number of steps of history does not affect the correlation a great deal, an insight that is reinforced when comparing the patterns of these measures for the vertices of interest in the 500m Soho graph in Figure 32.

Although RGI and RGIU do not seem closely related to the graph centrality measures for the grid world, the much closer correlation for the vertices of interest in the Soho street network provides confidence that they may correlate with brain activity as reported in Javadi et al. (2017).



(a) RGI $\beta = 10$ (b) RGIU $\beta = 10, n_{steps} = 1$ (c) RGIU $\beta = 10, n_{steps} = 2$

(d) Degree Centrality (e) Closeness Centrality (f) Betweenness Centrality

Figure 32: RGI $\beta = 10$ and RGIU $\beta = 100, n_{steps} \in \{1, 2\}$ for the Soho street network at 500m radius, compared with Degree Centrality, Closeness Centrality and Betweenness Centrality. Only junctions (vertices) used in Javadi et al. (2017) are shown, other junctions and connecting streets (edges) are hidden. Colour ranges are specific to each graph. The correlation between RGIU and betweenness centrality gets stronger with additional steps of history. This is because with many steps of the past known, in most states the direction of the goal becomes clearer. However at states which join less connected zones, such as the doorways in the six room scenario, knowledge of the history provides much less information about the location of the goal. These are also the states with highest betweenness centrality.

5.5 Correlation of information-theoretic measures with Space Syntax Measures

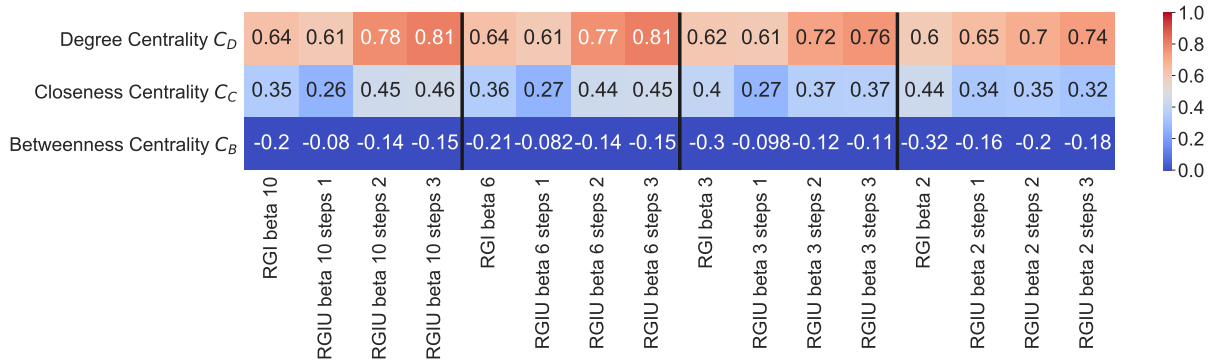


Figure 33: Correlation with $\beta \in \{10, 6, 3, 2\}$ of RGI, and RGIU $n_{steps} \in [1, 3]$ with Space Syntax segment averaged measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the vertices of interest ($n = 31$) in the Soho street network at 500m radius. RGI and RGIU at all values of β correlate with degree centrality.

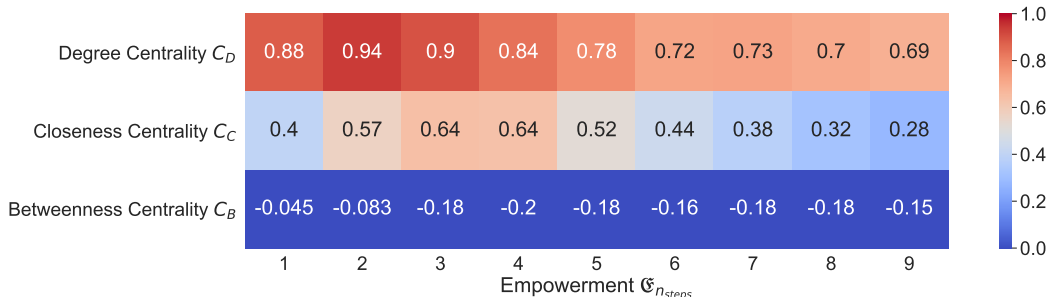


Figure 34: Correlation of Empowerment $\mathfrak{E}_{n_{steps} \in [1,9]}$ with Space Syntax segment averaged measures: degree centrality C_D, closeness centrality C_C and betweenness centrality C_B for the vertices of interest ($n = 31$) in the Soho street network at 500m radius. Empowerment is strongly correlated with degree centrality, and correlated with closeness centrality.

The Space Syntax measures reported in Javadi et al. (2017) were Degree Centrality C_D ("Connectivity" in Space Syntax terminology), Closeness Centrality C_C ("Integration") and Betweenness Centrality C_B ("Choice"). These are "segment" (edge) measures, but the data includes "Junction averages" which are averages of the segments connected to each junction.

Figure 33 and Figure 34 show the correlation of empowerment, RGI and RGIU with the Space Syntax junction average centrality measures for the 31 vertices of interest. Surprisingly betweenness centrality is slightly negatively correlated with any of the information-theoretic measures. The reasons for this are unclear as the details of the method of computation of the Space Syntax measures are not available, however it may be because betweenness centrality was computed for the whole of London, rather than for just a radius around Soho. Closeness centrality is fairly well correlated with empowerment with a maximum of Pearson $r = 0.64$ at 3 and 4 steps, falling to $r = 0.4$ at 1 step and $r = 0.28$ at 9 steps. Degree centrality is very strongly correlated with empowerment, with a maximum of $r = 0.94$ at 2 steps falling to $r = 0.88$ at 1 step and $r = 0.69$

at 9 steps. RGI is fairly well correlated with degree centrality $r = 0.64$ at $\beta = 10$ and RGIU is strongly correlated with a maximum of $r = 0.81$ at $\beta = 10$ and $n_{steps} = 3$.

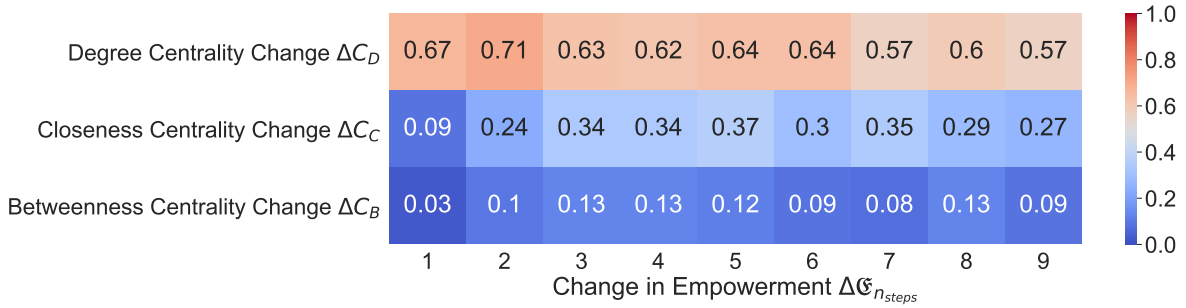


Figure 35: Correlation of Change in Empowerment $\Delta\mathfrak{E}_{n_{steps}\in[1,9]}$ with Space Syntax measures: change in degree centrality ΔC_D, change in closeness centrality ΔC_C and change in betweenness centrality ΔC_B for the vertices of interest ($n = 31$) in the Soho street network at 500m radius. The change in empowerment is correlated with the change in degree centrality.

The main results of Javadi et al. (2017) were correlations of brain activity with the *change* in street segment (edge) centrality (ΔC) on street entry after crossing a junction. The change in Empowerment ($\Delta\mathfrak{E}$) computed as the difference between the empowerment of the current junction and the previous junction. Where v_t is the current junction (vertex), v_{t-1} is the previous junction (vertex), e_t is the street segment (edge) being entered, e_{t-1} is the previous street segment (edge), $\Delta C_t = C(e_t) - C(e_{t-1}), \Delta\mathfrak{E}_t = \mathfrak{E}(v_t) - \mathfrak{E}(v_{t-1})$. RGI and RGIU do not show any significant correlation with the change of Degree Centrality, the change in Closeness Centrality, or the change in Betweenness Centrality. However, unsurprisingly, the change in Empowerment is correlated with the change in Degree Centrality as shown in Figure 35. The change in these measures depend on the route taken by the agent through the network, and were computed from the routes taken by participants during the Soho experiment. An agent with empowerment-maximising policy would need to track empowerment in the brain, or possibly to track changes in empowerment. Both empowerment and the change in empowerment will be good candidates for correlation when comparing with fMRI data in future work.

6 Discussion and Future Work

6.1 Why does the hippocampus appear to track centrality?

(Javadi et al., 2017) showed that activation in the right posterior hippocampus tracks the change in the degree centrality of the streets when the participants exited a junction in the virtual route through Soho. But why is tracking degree centrality useful to humans navigating the city? To answer this question I suggest that what may actually be being tracked is not the change in centrality per se, but the change in empowerment caused by a transition from one state (street) to another. Empowerment in a simple discrete model such as the one I employ here, is closely mathematically related to degree centrality, but empowerment is a deeper concept, rooted in the treatment of the perception-action loop as a communication channel, that can be applied in continuous and stochastic domains more like the natural environment where the ability to navigate evolved. Perhaps we have learnt to stay empowered, all else being equal, and this manifests itself in the discretised navigational context of the city as tracking the change in degree centrality.

6.2 The potential for empowerment and relevant information models for planning the built environment

Empowerment has enjoyed considerable success as a pseudo-utility for agents that induces intelligent like behaviour in a range of modelled scenarios (Klyubin et al., 2005*b*,*a*, 2008; Capdepuy et al., 2007; Polani, 2009; Capdepuy et al., 2012; Salge et al., 2014*b*; Mohamed and Rezende, 2015; Karl et al., 2015, 2017; Clements and Polani, 2017), and also in robots (Catenacci Volpi et al., 2016). The theory that humans in certain situations are intrinsically motivated by empowerment has not been tested empirically. The correlation of empowerment to the graph-theoretic measures in the familiar six room grid world and the Soho street network provides a first step towards determining if empowerment does indeed predict human behaviour in navigation, the earliest of decision-making domains. Informational modelling of decision making under cognitive constraints, termed "bounded rationality" after the work by Simon (1957) has recently gained momentum as a research topic (e.g. Gottwald and Braun (2019)). Relevant goal information, and relevant goal information uptake offer a perspective on the structure of the environment that encodes information processing cost and memory constraints. The optimisation of the built environment for productivity, and to minimising energy consumption, is a hot topic worldwide encapsulated in the mot-du-jour "Smart Cities". With more accurate predictions of the flow and occupancy of buildings, architects can optimise designs, building managers can optimise usage, and automated systems controlling heating and air-conditioning can avoid wasting energy heating or cooling a space that is unlikely to be used. Improved models of vehicular, bicycle and pedestrian traffic offer opportunities for city planners to reduce pollution, reduce journey times and promote healthy transportation modes. Agent-based models utilising Empowerment, RGI and RGIU offer researchers hoping to predict human behaviour new tools that complement the existing methods provided by Space Syntax, OSMNX (Boeing, 2017) and others. Distilling the software tools developed here into an open source library that could be integrated with OSMNX or DepthMapX(depthmapX development team, 2017) would encourage other researchers to explore the insights offered by the measures.

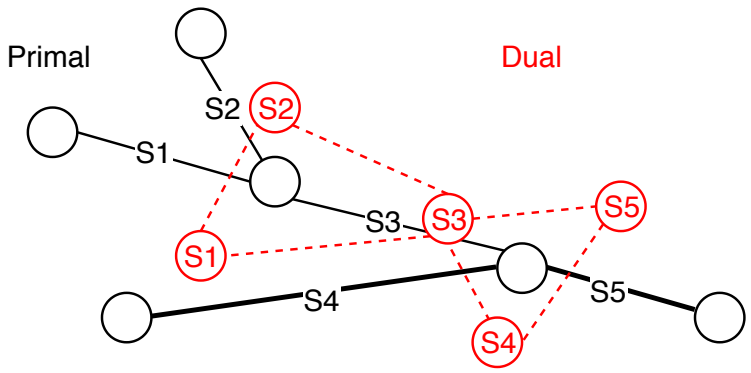


Figure 36: Primal and dual graphs of a small street network. Street segments are labelled S1, S2, S3, S4, S5. The dual graph, known as the "information space" can be simplified by the removal of vertices from considering street segments to be part of the same road. If S4 and S5 are both wider streets they could be considered to be a single road. Alternatively S1,S2 and S3 could be a single road because the angle between them is very small.

6.3 The "information space" of the dual graph

The primal graph of the street network is a planar graph embedded in Euclidean space, where junctions are vertices and street segments are edges. A dual graph generated from a street network map, described as the "information space" (Masucci et al., 2014), has a vertex for each street segment, and an edge connecting vertices which are street segments that are connected by a junction. In graph theory this special kind of dual graph is known as a *line graph*, and it has been proved that for all but one exception, connected graphs can be generated completely from their complementary line graph (Whitney, 1932). Figure 36 shows the primal and dual graph of a small street network. The dual graph can be simplified in various ways: by angle of the streets, by size or capacity or based on the naming of streets (e.g. Regent street in London has multiple segments but it is still one street) (Jiang et al., 2008; Masucci et al., 2014). This compression of the state space in the dual graph has informational benefits – it is much easier to say "Follow Regent Street until you reach Piccadilly", than a series of directions for each junction along the way. The naming of streets could also be thought of as a result of this compression rather than a source. Why are the individual segments of Regent street grouped together and given a single name? Whether we look at the width of the road, the amount of traffic, the type of shops, or the angle of the segments, the naming of the road is an indication of the human drive to simplify the representation of space, because it confers informational advantages that once again can be attributed to the need for information parsimony in a complex world. The derivation of the dual representation of the primal graph for the street network is a deep and rich topic that has been explored empirically (Jiang et al., 2008; Masucci et al., 2009, 2014) but not from an information-theoretic perspective based on the biological necessity of information parsimony.

Empowerment and relevant information also provide ways of automatically compressing the state space by identifying states where the flow of information between senses and actions is likely to be high, or where the amount of information that needs to be loaded or unloaded is high (Polani, 2009; Van Dijk and Polani, 2011; Anthony et al., 2011). The clustering of states into action sequences or "options" (Sutton and Singh, 2015) has been shown to confer informational benefits (McNamee et al., 2016), and indeed modelling the hierarchical representation of actions and states

achieved by the human brain is the "holy grail of artificial intelligence research" (Catenacci Volpi, 2019). The concept of spatial hierarchies has large body of previous work (e.g. Kuipers (2008)), but the informational measures presented here may be able to shed some new light on how these semantic hierarchies are constructed from the perspective of cognitive constraints. With this in mind, an informational treatment of the dual graph for street networks seems like a rich seam of research opportunity.

The results in Section 5 for empowerment and relevant goal information are computed on the primal graph. Given the practical success of the application of both the primal and dual graph in spatial analysis, an approach which encapsulates the advantages of both is enticing, and some work has been done in this direction with a Markov chain analysis by Batty (2017). The adaptation of relevant information and empowerment to a dual graph representation, (or a mixed model) will be an exciting avenue for future work.

6.4 Possible extensions to relevant goal information uptake

6.4.1 Reuse of actions

As discussed in Section 4.3 relevant goal information does not convey the same benefit of compressing the state space as relevant information – a minimum RGIU policy does not favour the reuse of actions across states. RGIU at first glance would not seem to be a good candidate for expressing the informational saving of being able to say when giving directions in the city "Keep going straight on until you reach the T-junction". There may be a way of deriving this benefit from another look at how the policy is learnt by the agent, or by considering the dual graph.

6.4.2 I^{old}

RGIU is an estimation of the new amount of information needed about the goal, in a particular state, given the full history of the agent dubbed I^{new} (Van Dijk and Polani, 2013). The same paper also introduces I^{old} as the information no longer required about the goal in a particular state, which can be dropped from working memory. It is not clear whether brain activation can be expected to show information entering or leaving working memory, or both.

6.5 Connecting empowerment and relevant information with current research into active inference in the context of studies of the neuroscience of mammalian navigation

The role of the hippocampus in navigation, but also in generalised decision making underlines not only the spectacular ability of the mammalian brain to reuse neural architecture across multiple modalities but also the role of navigation as the earliest of decision making domains. During the Soho navigation experiment subjects were periodically "surprised" at junctions where the optimal route was closed to them. Rigoli et al. (2019) report that the hippocampus has a role in modulating response to prediction errors during inference. An animal operating in the real world is constantly faced by the challenge of navigating a changing environment. With this in mind it is perhaps unsurprising that given the role of the hippocampus in tracking aspects of the topology (Javadi et al., 2017), and the animals place in the world (Howard et al., 2014), that it also has a role in minimising surprise. Too much surprise in the animals model of the environment could lead to falling off a cliff or being consumed by a predator. The minimisation of surprise, in other

words the minimisation of error between the agent's predictive model and the environment, as a driver for behaviour is a central theme of active inference. Empowerment and relevant information are not restricted to fully-observable, deterministic transition models, the formalism can be naturally extended to partially observable or changing environments. Under this regime maximising empowerment is an effective way to minimise surprise in terms of the agent's ability to influence the world state in the future, by acting upon itself (e.g. movement) or by changing the environment (e.g. digging). The mathematical relationship between empowerment and free energy (underpinning active inference) is explored in Biehl et al. (2015, 2018), showing deep similarities in the formalisms. Future work which brings together the growing evidence base for active inference, with empowerment, relevant information and associated work, in the context of empirical studies of animal navigational behaviour is an exciting prospect.

6.6 Comparison with empirical neuroscience

This work has been inspired by a collaboration, originally conceived by my supervisor Daniel Polani and Hugo Spiers, to investigate the possibility that information-theoretic measures of the topology of the environment can predict navigational behaviour in humans and other mammals, via a comparison with neural activity. I have argued that empowerment, relevant goal information, and relevant goal information uptake uncover new aspects of the topology of street networks, when compared to existing metrics based on graph centrality. These information-theoretic quantities are founded on an established evolutionary biological principle of information parsimony and the model of the perception-action loop as a communication channel. The correlation of the information-theoretic measures with existing graph-theoretic measures suggest that a direct comparison with human brain activation and behaviour may be a fruitful avenue of scientific exploration.

A minimum RGI policy demonstrates an agent that is parsimonious in terms of information needed to operate, and RGIU shows how working memory constraints can be modelled. The interplay between these two constraints has not been investigated in this work, and the tradeoff that increasing the empowerment horizon must entail is similarly not yet explored. The increasing temporal and spatial resolution of brain activation, alongside the rapid progress in computational neuroscience offers an exciting opportunity to test more complex theoretical models of cognitive constraints with empirical data.

References

Albani, A. E., Mangano, M. G., Buatois, L. A., Bengtson, S., Riboulleau, A., Bekker, A., Konhauser, K., Lyons, T., Rollion-Bard, C., Bankole, O., Baghekema, S. G. L., Meunier, A., Trentesaux, A., Mazurier, A., Aubineau, J., Laforest, C., Fontaine, C., Recourt, P., Fru, E. C., Macchiarelli, R., Reynaud, J. Y., Gauthier-Lafaye, F. and Canfield, D. E. (2019), 'Organism motility in an oxygenated shallow-marine environment 2.1 billion years ago', *Proc. Natl. Acad. Sci.* **116**(9), 3431–3436.
doi:10.1073/PNAS.1815721116

Al_Sayed, K., Turner, A., Hillier, B. and Iida, S. (2014), *Space Syntax Methodology*, 4th edn, Bartlett School of Architecture, UCL, London.

Anthony, T., Polani, D. and Nehaniv, C. L. (2011), Impoverished empowerment: 'meaningful' action sequence generation through bandwidth limitation, *in* G. Kampis, I. Karsai and E. Szathmáry, eds, 'Advances in Artificial Life. Darwin Meets von Neumann', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 294–301.

Arimoto, S. (1972), 'An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels', *IEEE Trans. Inf. Theory* **18**(1), 14–20.
URL: *http://ieeexplore.ieee.org.ezproxy.herts.ac.uk/stamp/stamp.jsp?arnumber=1054753*

Ashby, R. W. (1956), *An Introduction to Cybernetics*, Chapman & Hall LTD, London.
URL: *http://pcp.vub.ac.be/books/IntroCyb.pdf*

Ay, N., Bertschinger, N., Der, R., Güttler, F. and Olbrich, E. (2008), 'Predictive information and explorative behavior of autonomous robots', *The European Physical Journal B* **63**(3), 329–339.

Batty, M. (2017), Space Syntax and Spatial Interaction: Comparisons, Integrations, Applications, Technical Report February, Bartlett School of Architecture, UCL.
doi:10.13140/RG.2.2.15233.25447

Bellman, R. (1953), An introduction to the theory of dynamic programming, Technical report, Rand Corporation, Santa Monica CA, USA.

Bellman, R. (1957), 'A markovian decision process', *Indiana Univ. Math. J.* **6**, 679–684.

Bertsekas, D. P. (2005), *Dynamic Programming and Optimal Control*, Vol. I, 3rd edn, Athena Scientific, Belmont, MA, USA.

Bialek, W. and Tishby, N. (1999), Predictive Information, Technical report, arXiv preprint cond-mat/9902341.
URL: *https://arxiv.org/pdf/cond-mat/9902341.pdf*

Biehl, M., Guckelsberger, C., Salge, C., Smith, C. and Polani, D. (2015), Free energy, empowerment, and predictive information compared, Technical report, University of Hertfordshire.
URL: *https://www.mis.mpg.de/fileadmin/pdf/abstract_gso18_3300.pdf*

Biehl, M., Guckelsberger, C., Salge, C., Smith, S. C. and Polani, D. (2018), 'Expanding the active inference landscape: More intrinsic motivations in the perception-action loop', *Front. Neurorobot.* **12**(AUG).
doi:10.3389/fnbot.2018.00045

Blahut, R. (1972), 'Computation of channel capacity and rate-distortion functions', *IEEE transactions on Information Theory* **18**(4), 460–473.

Boeing, G. (2017), 'OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Comput. Environ. Urban Syst.* **65**, 126–139.
doi:10.1016/j.compenvurbsys.2017.05.004

Botvinick, M. M., Niv, Y. and Barto, A. C. (1995), 'Hierarchically organized behavior and its neural foundations: A reinforcement-learning perspective', *Cognition* **113**(3), 262–280.
doi:10.1016/j.cognition.2008.08.011

Brandes, U. and Erlebach, T. (1998), Centrality Indices, *in* 'LNCS 3418 - Netw. Anal.', Springer, chapter 3, pp. 16–60.
URL: *https://link.springer.com/content/pdf/10.1007%2Fb106453.pdf*

Capdepuy, P., Polani, D. and Nehaniv, C. L. (2007), 'Maximization of potential information flow as a universal utility for collective behaviour', *Proc. 2007 IEEE Symp. Artif. Life, CI-ALife 2007* pp. 207–213.
doi:10.1109/ALIFE.2007.367798

Capdepuy, P., Polani, D. and Nehaniv, C. L. (2012), 'Perception-Action Loops of Multiple Agents: Informational Aspects and the Impact of Coordination', *Theory Biol. Sci.* **131**(3), 149–159.
URL: *http://uhra.herts.ac.uk/bitstream/handle/2299/8997/906076.pdf?sequence=1*

Catenacci Volpi, N. (2019), personal communication.

Catenacci Volpi, N., De Palma, D., Polani, D. and Indiveri, G. (2016), 'Computation of Empowerment for an Autonomous Underwater Vehicle', *IFAC-PapersOnLine* **49**(15), 81–87.
doi:10.1016/j.ifacol.2016.07.713

Cheng, Q. and Chen, F. (2013), 'Variational Planning for Graph-based MDPs', *Nips* pp. 1–9.
URL: *https://papers.nips.cc/paper/5067-variational-planning-for-graph-based-mdps.pdf*

Clark, A. (2013), 'Whatever next? predictive brains, situated agents, and the future of cognitive science', *Behavioral and brain sciences* **36**(3), 181–204.

Clements, M. and Polani, D. (2017), Empowerment as a Generic Utility Function for Agents in a Simple Team Sport Simulation, *in* 'Interact. Collab. Robot. - Second Int. Conf. 2017, Hatfield, UK, Sept. 12-16, 2017, Proc.', Springer, Cham, pp. 37–49.
doi:10.1007/978-3-319-66471-2_5

Cover, T. M. and Thomas, J. A. (2006), *Elements of Information Theory*, 2nd edn, Wiley.
doi:10.1007/978-94-010-9292-0

Darwin, C. (1859), *On the Origin of Species by Means of Natural Selection*, Murray, London. or the Preservation of Favored Races in the Struggle for Life.

Dayan, P., Hinton, G. E., Neal, R. M. and Zemel, R. S. (1995), 'The Helmholtz machine.', *Neural Comput.* **7**(5), 889–904.
doi:10.1162/neco.1995.7.5.889

depthmapX development team (2017), 'depthmapx'.
URL: *https://github.com/SpaceGroupUCL/depthmapX/*

Epstein, R. A., Patai, E. Z., Julian, J. B. and Spiers, H. J. (2017), 'The cognitive map in humans: Spatial navigation and beyond'.

Freeman, L. C. (1977), 'A Set of Measures of Centrality Based on Betweenness', *Sociometry* **40**(1), 35.
doi:10.2307/3033543

Friston, K. (2009), 'The free-energy principle: a rough guide to the brain?', *Trends Cogn. Sci.* **13**(7), 293–301.
doi:10.1016/j.tics.2009.04.005

Friston, K. (2010), 'The free-energy principle: a unified brain theory?', *Nature reviews neuroscience* **11**(2), 127.

Friston, K., Kilner, J. and Harrison, L. (2006), 'A free energy principle for the brain', *Journal of Physiology-Paris* **100**(1-3), 70–87.

Fuster, J. M. (1990), 'Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle', *Annals of the New York Academy of Sciences* **608**(1), 318–336.
doi:10.1111/j.1749-6632.1990.tb48901.x

Garvert, M. M., Dolan, R. J. and Behrens, T. E. (2017), 'A map of abstract relational knowledge in the human hippocampal-entorhinal cortex', *Elife* **6**, 1–20.
doi:10.7554/eLife.17086

Gottwald, S. and Braun, D. A. (2019), 'Bounded rational decision-making from elementary computations that reduce uncertainty', *Entropy* **21**(4).
doi:10.3390/e21040375

Grinstead, C. M. and Snell, J. L. (2006), Markov Chains (Chapter 11), *in* 'Introd. to Probab.', University Press of Florida, chapter 11, pp. 405–470.

Hagberg, A., Swart, P. and S Chult, D. (2008), Exploring network structure, dynamics, and function using networkx, Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Hartley, T., Maguire, E. A., Spiers, H. J. and Burgess, N. (2003), 'The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans', *Neuron* **37**(5), 877–888.
doi:10.1016/S0896-6273(03)00095-3

Hiller, B. and Iida, S. (2005), Network effects and psychological effects: a theory of urban movement, *in* 'Spat. Inf. Theory', pp. 475–490.
URL: *http://spacesyntax.tudelft.nl/media/Long papers I/hillieriida.pdf*

Hillier, B. and Hanson, J. (1984), *The Social Logic of Space*, Cambridge University Press.
doi:10.1017/CBO9780511597237

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., Loftus, M. M., Staskute, L. and Spiers, H. J. (2014), 'The hippocampus and entorhinal cortex encode the path and euclidean distances to goals during navigation', *Curr. Biol.* **24**(12), 1331–1340. doi:10.1016/j.cub.2014.05.001

Howard, R. A. (1960), *Dynamic programming and markov processes.*, John Wiley.

Howard, R. A. (1966), 'Information Value Theory', *Syst. Sci. Cybern. IEEE Trans.* **2**(1), 22–26. doi:10.1109/TSSC.1966.300074

Howard, R. A. and Matheson, J. E. (1984), Influence diagrams, *in* R. A. Howard and J. E. Matheson, eds, 'The Principles and Applications of Decision Analysis, Volume II', Microsoft, Strategic Decisions Group, Menlo Park, CA.

James, R. G., Ellison, C. J. and Crutchfield, J. P. (2018), 'dit: a Python package for discrete information theory', *The Journal of Open Source Software* **3**(25), 738. doi:https://doi.org/10.21105/joss.00738

Javadi, A.-H., Emo, B., Howard, L. R., Zisch, F. E., Yu, Y., Knight, R., Pinelo Silva, J. and Spiers, H. J. (2017), 'Hippocampal and prefrontal processing of network topology to simulate the future', *Nat. Commun.* **8**, 14652. doi:10.1038/ncomms14652

Jaynes, E. T. (1957*a*), 'Information theory and statistical mechanics', *Physical review* **106**(4), 620.

Jaynes, E. T. (1957*b*), 'Information theory and statistical mechanics. ii', *Physical review* **108**(2), 171.

Jiang, B., Zhao, S. and Yin, J. (2008), 'Self-organized Natural Roads for Predicting Traffic Flow: A Sensitivity Study', *J. Stat. Mech. Theory Exp.* . **URL:** *https://arxiv.org/pdf/0804.1630.pdf*

Jung, T., Polani, D. and Stone, P. (2011), 'Empowerment for continuous agent-environment systems', *Adaptive Behavior* **19**(1), 16–39.

Kahneman, D. (2011), *Thinking, fast and slow.*, Farrar, Straus and Giroux, New York, NY, US.

Kahneman, D., Slovic, S. P., Slovic, P. and Tversky, A. (1982), *Judgment under uncertainty: Heuristics and biases*, Cambridge university press.

Kahneman, D. and Tversky, A. (1979), 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica* **47**(2), 263–292. **URL:** *http://www.its.caltech.edu/ camerer/Ec101/ProspectTheory.pdf*

Kandel, E., Schwartz, J. and Jessell, T. (1991), *Principles of Neural Science*, 3rd edn, McGraw-Hill, NewYork.

Karl, M., Bayer, J. and van der Smagt, P. (2015), Efficient Empowerment, Technical report, Technische Universitat Munchen, Munich.

Karl, M., Soelch, M., Becker-Ehmck, P., Benbouzid, D., van der Smagt, P. and Bayer, J. (2017), 'Unsupervised Real-Time Control through Variational Empowerment'.
URL: *https://arxiv.org/pdf/1710.05101.pdf http://arxiv.org/abs/1710.05101*

Klyubin, A., Polani, D. and Nehaniv, C. (2005*a*), 'Empowerment: A universal agent-centric measure of control', *Evol. Comput. 2005 IEEE Congr.* **1**, 128–135.

Klyubin, A. S., Polani, D. and Nehaniv, C. L. (2005*b*), All Else Being Equal Be Empowered, *in* 'Adv. Artif. Life', Springer, pp. 744–753.
doi:10.1007/11553090_75

Klyubin, A. S., Polani, D. and Nehaniv, C. L. (2008), 'Keep Your Options Open: An Information-Based Driving Principle for Sensorimotor Systems', *PLoS One* **3**(12), e4018.
doi:10.1371/journal.pone.0004018

Klyubin, A. S., Polani, D., Nehaniv, C. L., Lane, C., Kljubin, E.-m. A., Polani, D. and Nehaniv, C. L. (2004), Organization of the Information Flow in the Perception-Action Loop of Evolved Agents, *in* 'Evolvable Hardware. Proceedings. 2004 NASA/DoD Conf.', IEEE, Seattle, WA, USA, pp. 177–180.
doi:10.1109/EH.2004.1310828

Kool, W., Mcguire, J. T. and Botvinick, M. M. (2010), 'Decision Making and the Avoidance of Cognitive Demand', *J. Exp. Psychol. Gen.* .
doi:10.1037/a0020198

Kuipers, B. (2008), An intellectual history of the spatial semantic hierarchy, *in* M. E. Jefferies and W.-K. Yeap, eds, 'Robotics and Cognitive Approaches to Spatial Mapping', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 243–264.

Lee, D., Seo, H. and Jung, M. W. (2012), 'Neural basis of reinforcement learning and decision making.', *Annu. Rev. Neurosci.* **35**, 287–308.
doi:10.1146/annurev-neuro-062111-150512

Masucci, A. P., Smith, D., Crooks, A. and Batty, M. (2009), 'Random planar graphs and the London street network', *Eur. Phys. J. B* **71**(71), 259–271.
doi:10.1140/epjb/e2009-00290-4

Masucci, A. P., Stanilov, K. and Batty, M. (2014), 'Exploring the Evolution of London's Street Network in the Information Space: a Dual Approach', *Phys. Rev. E* **89**(1).
doi:10.1103/PhysRevE.89.012805

McCulloch, W. and Pitts, W. (1943), 'A logical calculus of ideas immanent in nervous activity', *Bulletin of Mathematical Biophysics* **5**, 127–147.

McNamee, D., Wolpert, D. M. and Lengyel, M. (2016), Efficient state-space modularization for planning: theory, behavioral and neural signatures, *in* D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds, 'Advances in Neural Information Processing Systems 29', Curran Associates, Inc., pp. 4511–4519.
URL: *http://papers.nips.cc/paper/6320-efficient-state-space-modularization-for-planning-theory-behavioral-and-neural-signatures.pdf*

Mohamed, S. and Rezende, D. J. (2015), Variational information maximisation for intrinsically motivated reinforcement learning, *in* 'Adv. Neural Inf. Process. Syst.', Vol. 2015-Janua, pp. 2125–2133.

O'Keefe, J. and Dostrovsky, J. (1971), 'The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat.', *Brain research* .

O'Keefe, J. and Nadel, L. (1978), *The hippocampus as a cognitive map*, Oxford: Clarendon Press.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Penny, W. (2012), 'Bayesian models of brain and behaviour', *ISRN Biomathematics* **2012**.

Penny, W. D., Zeidman, P. and Burgess, N. (2013), 'Forward and Backward Inference in Spatial Cognition', *PLoS Comput. Biol.* **9**(12).
doi:10.1371/journal.pcbi.1003383

Polani, D. (2009), 'Information: Currency of life?', *HFSP J.* **3**(5), 307–316.
doi:10.2976/1.3171566

Polani, D. (2011), An informational perspective on how the embodiment can relieve cognitive burden, *in* '2011 IEEE Symposium on Artificial Life (ALIFE)', IEEE, pp. 78–85.

Polani, D., Martinetz, T. and Kim, J. (2001), An Information-Theoretic Approach for the Quantification of Relevance, *in* J. Kelemen and P. Sosík, eds, 'Lect. Notes Comput. Sci.', Vol. 2159, Springer.
URL: *https://link.springer.com/content/pdf/10.1007/3-540-44811-X_82.pdf*

Polani, D., Nehaniv, C., Martinetz, T. and Kim, J. (2006), 'Relevant information in optimized persistence vs. progeny strategies', *Proc. Artificial Life X* .

Polani, D., Sporns, O. and Lungarella, M. (2007*a*), 'How Information and Embodiment Shape Intelligent Information Processing', *50 Years Artif. Intell.* pp. 99–111.
doi:10.1007/978-3-540-77296-5_10

Polani, D., Sporns, O. and Lungarella, M. (2007*b*), 'How Information and Embodiment Shape Intelligent Information Processing', *50 Years Artif. Intell.* pp. 99–111.
doi:10.1007/978-3-540-77296-5_10

Porta, S., Crucitti, P. and Latora, V. (2006), 'The network analysis of urban streets: A primal approach', *Environ. Plan. B Plan. Des.* **33**(5), 705–725.
doi:10.1068/b32045

Powers, W. T. (1973), *The control of perception*, Aldine, Oxford, England.

Rescorla, M. (2015), 'The Computational Theory of Mind'.
URL: *https://plato.stanford.edu/entries/computational-mind/*

Rigoli, F., Michely, J., Friston, K. J. and Dolan, R. J. (2019), 'The role of the hippocampus in weighting expectations during inference under uncertainty', *Cortex* **115**, 1–14.
doi:10.1016/j.cortex.2019.01.005

Russell, S. J. and Norvig, P. (2009), *Artificial intelligence: a modern approach*, third edn, Prentice Hall.
doi:10.1049/me:19950308

Salge, C., Glackin, C. and Polani, D. (2013), 'Approximation of empowerment in the continuous domain', *Advances in Complex Systems* **16**(02n03), 1250079.

Salge, C., Glackin, C. and Polani, D. (2014*a*), 'Changing the environment based on empowerment as intrinsic motivation', *Entropy* **16**(5), 2789–2819.

Salge, C., Glackin, C. and Polani, D. (2014*b*), Empowerment – An Introduction, *in* M. Prokopenko, ed., 'Guid. Self-Organization Inception', Springer.

Salge, C. and Polani, D. (2017), 'Empowerment as replacement for the three laws of robotics', *Frontiers in Robotics and AI* **4**, 25.

Schmidhuber, J. (2010), 'Formal theory of creativity, fun, and intrinsic motivation (1990-2010)', *Auton. Ment. Dev. IEEE Trans.* **2**(3), 230–247.

Seth, A. K. (2013), 'Interoceptive inference, emotion, and the embodied self', *Trends in cognitive sciences* **17**(11), 565–573.

Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *Bell Syst. Tech. J.* **27**(4), 623–656.
doi:10.1002/j.1538-7305.1948.tb00917.x

Simon, H. (1957), 'A behavioral model of rational choice, in models of man, social and rational: mathematical essays on rational human behavior in a social setting', *New York: Wyley* .

Simon, H. (1978), 'Rational Decision-Making in Business Organizations', *Am. Econ. Rev.* pp. 493–513.
doi:10.2307/1808698

Solway, A. and Botvinick, M. M. (2012), 'Goal-Directed Decision Making as Probabilistic Inference: A Computational Framework and Potential Neural Correlates', *Psychol. Rev.* **119**(1), 120–154.
doi:10.1037/a0026435.supp

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J. and Friston, K. J. (2005), 'Information theory, novelty and hippocampal responses: Unpredicted or unpredictable?', *Neural Networks* **18**(3), 225–230.
doi:10.1016/j.neunet.2004.12.004

Sutton, R. S. and Barto, A. G. (1998), *Introduction to Reinforcement Learning*, 1st edn, MIT Press, Cambridge, MA, USA.

Sutton, R. S. and Singh, S. (2015), 'Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning', *Artif. Intell.* **1**(1999), 181–211.
doi:10.1017/CBO9781107415324.004

Thornton, C. (2014), 'Infotropism as the underlying principle of perceptual organization', *Journal of Mathematical Psychology* **61**, 38–44.

Tolman, E. C. (1948), 'Cognitive Maps in Rats and Men', *Psychol. Rev.* .
URL: *https://pdfs.semanticscholar.org/e84d/660ebf894e2fb85d7b27985bfbf07b9acd22.pdf*

Touchette, H. and Lloyd, S. (2000), 'Information-Theoretic limits of control', *Phys. Rev. Lett.* **84**(6), 1156–1159.
doi:10.1103/PhysRevLett.84.1156

Van Dijk, S. G. (2013), Informational Constraints and Organisation of Behaviour, Phd thesis, University of Hertfordshire.

Van Dijk, S. G. and Polani, D. (2011), 'Grounding subgoals in information transitions', *IEEE SSCI 2011 Symp. Ser. Comput. Intell. - ADPRL 2011 2011 IEEE Symp. Adapt. Dyn. Program. Reinf. Learn.* pp. 105–111.
doi:10.1109/ADPRL.2011.5967384

Van Dijk, S. G. and Polani, D. (2012), 'Informational Organization of Task Structure'.

Van Dijk, S. G. and Polani, D. (2013), 'Informational constraints-driven organization in goal-directed behavior', *Advances in Complex Systems* **0**(0).
doi:10.1142/S0219525913500161

Van Dijk, S. G., Polani, D. and Nehaniv, C. L. (2010), 'What do You Want to do Today ? Relevant-Information Bookkeeping in Goal-Oriented Behaviour', *Artif. Life XII 12th Int. Conf. Synth. Simul. Living Syst.* pp. 176–183.
URL: *http://homepages.stca.herts.ac.uk/ sv08aav/pub/vandijk.alife10.goal.pdf*

Vergassola, M., Villermaux, E. and Shraiman, B. I. (2007), 'Infotaxis as a strategy for searching without gradients', *Nature* **445**(7126), 406.

Von Helmholtz, H. (1867), *Handbuch der physiologischen Optik*, Vol. 9, Voss.

Whitney, H. (1932), 'Congruent graphs and the connectivity of graphs', *American Journal of Mathematics* **54**(1), 150–168.
URL: *http://www.jstor.org/stable/2371086*

A Appendix: Supporting work

A.1 Relevant Goal Information Uptake weighted by $p(s_t)$

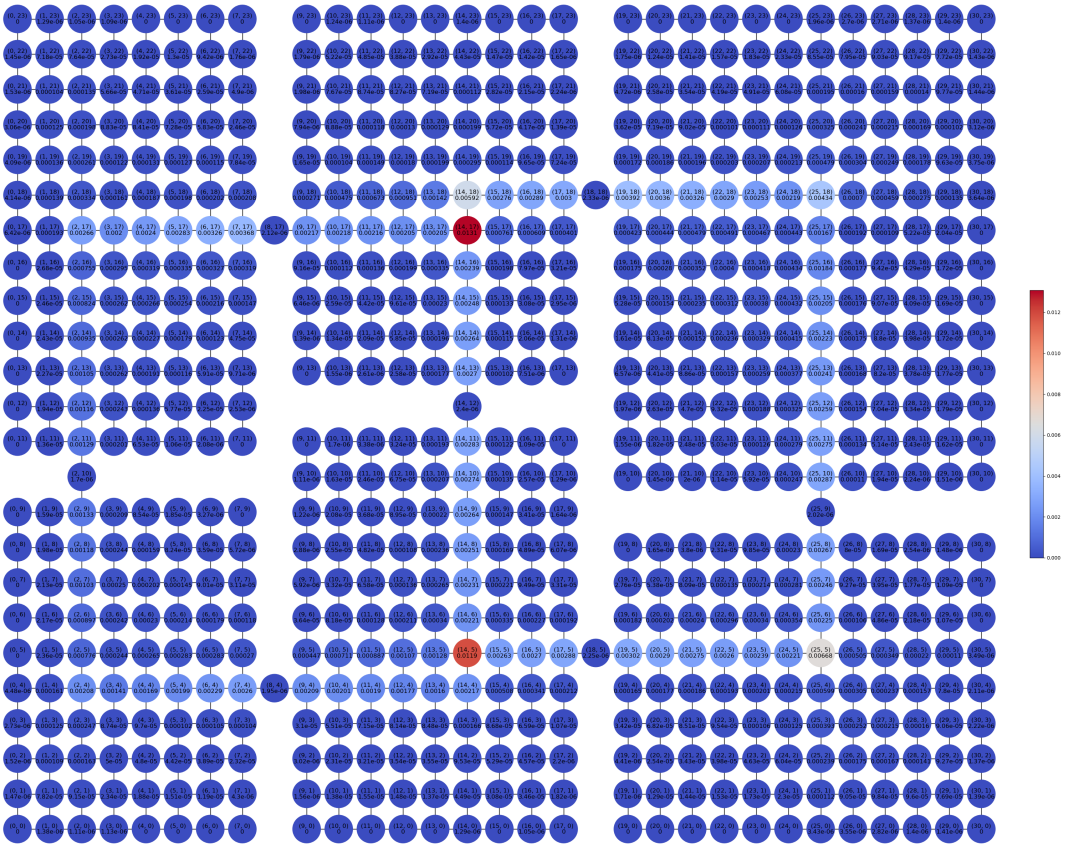


Figure 37: Relevant goal information uptake weighted by $p(s)$ (stationary distribution) for a 6 room grid world scenario. Although RGIU does rise in states close to the doorways, the effect is not as pronounced as demonstrated by Van Dijk, and the crossing points in the rooms opposite doorways are much more pronounced. The reasons for this are beyond the scope of this work to analyse, but could be due to the limited number of steps.