Are Psychological Variables Relevant to Evaluating Geoinformatics Applications? The Case of Landmarks (Vision Paper)

Jakub Krukar[1]
Institute for Geoinformatics, Universität Münster, Germany

Angela Schwering
Institute for Geoinformatics, Universität Münster, Germany

Abstract

Interdisciplinary integration of spatial cognition and spatial computation promises to create better spatial technology based on findings from cognitive psychology experiments. Using the example of psychological studies and computational modelling of landmarks, this paper argues that core evaluation criteria of both disciplines are not well aligned with the goal of evaluating landmark-enhanced navigation support systems that support users in everyday wayfinding.

The paper raises two points. First, it reviews evaluation criteria used in the interdisciplinary field of landmark research. It is argued that when to consider the role of landmark-enhanced navigation support systems in everyday life of their users, different evaluation criteria are needed. If strictly-psychological or strictly-computational criteria continue being prioritised by the community, we risk undervaluing significant technological contributions. Second, it proposes one such potential criterion: testing whether the cognitive task has changed due to equipping users with the new technology.

This goal might be achieved at the expense of criteria typical to strictly-psychological studies (such as spatial memory of landmarks along the travelled route) or strictly-computational studies (such as efficiency and accuracy of a landmark-selection algorithm). Thus, promoting and implementing alternative evaluation criteria comes with methodological risks. In order to mitigate them we propose a process based on pre-registration of "postdiction" studies and hope to stimulate a further debate on a consensus-based approach in the community.

2012 ACM Subject Classification Human-centered computing → Empirical studies in ubiquitous and mobile computing; Applied computing → Psychology

Keywords and phrases wayfinding, navigation support systems, cognitive geoengineering, landmarks

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.10

Category Vision Paper

Supplementary Material *Dataset (Literature Review)*: `https://osf.io/ru8yb/`

Funding This work has been supported by the German Research Foundation grant SCHW1372/7-3.

Acknowledgements We wish to thank Eva Nuhn and Sabine Timpf for organising the workshop on "Modelling of landmark dimensions" held at the 8th International Conference on Spatial Cognition 2021, as well as all its participant for an inspiring discussion that lead us to writing this paper.

1 Introduction

Research on landmarks can be classified into three main streams:

(a) *psychological research* aimed at understanding how the mind works (e.g., understanding the role of landmarks in spatial cognition);

[1] Corresponding author

15th International Conference on Spatial Information Theory (COSIT 2022).
Editors: Toru Ishikawa, Sara Irina Fabrikant, and Stephan Winter; Article No. 10; pp. 10:1–10:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

(b) *computational research* aimed at solving the problem of landmark identification based on human-specific criteria (e.g., modelling how the human mind would identify salient landmarks); and

(c) *navigation support research* that aims to support the user in everyday wayfinding by enhancing existing technology with landmarks.

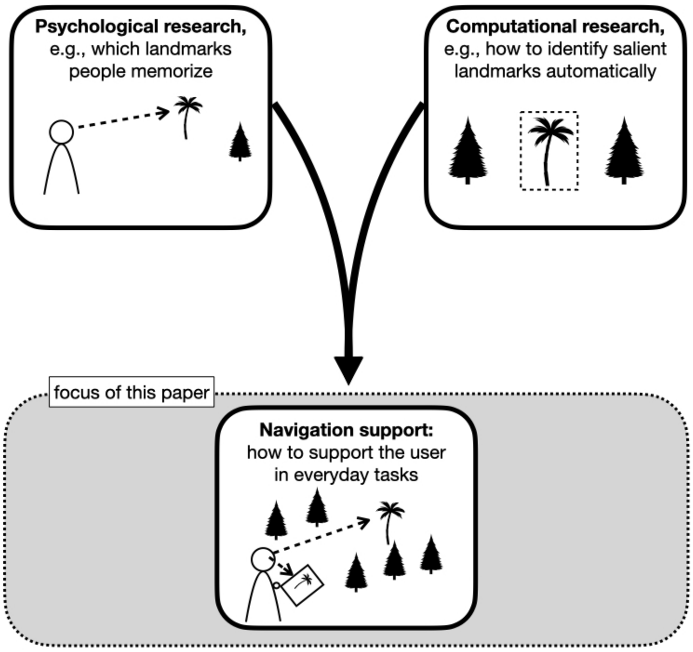

Figure 1 The practical goal of interdisciplinary collaboration in landmark research is the focus of this paper.

The practical aim of the community's interdisciplinary effort is to create (c) (Figure 1). Yet, this effort seems to have reached an impasse. Despite fundamental landmark studies spanning back at least six decades [13] it remains impossible to predict which specific landmark should be presented or highlighted for each individual user in each possible context of use; and it is uncertain whether it should be. The field's momentum seems to be directed at modelling more factors contributing to the context of use, identifying more dimensions contributing to personalised preferences, and extracting more types of potentially relevant digital data. Yet, there has been criticism of the entire concept of landmarks underlying this line of research. Montello [15] argued that landmarks are exaggerated in both basic and applied spatial cognition research, reviewing two issues in particular: that the term "landmark" is a label for several concepts; and that – even when communicated precisely – the concept of landmarks has an overestimated role in human navigation, compared to other components of spatial knowledge, memory, and reasoning.

The current paper focuses on another possibility underlying the impasse: that landmark-enhanced navigation support systems are evaluated based on criteria relevant to basic (be it psychological or computational) research, but with no clear agreement on which criteria should be employed for evaluating their applications. This paper reviews two pathways for moving beyond the impasse:

(1) re-considering the role of landmark-enhanced navigation support systems in everyday life of their users; and
(2) standardising the evaluation of a phenomenon difficult to capture with traditional measures – whether the technology changed the cognitive task performed by its user.

In order to do so, the paper presents a re-analysis of a recently published review of papers on landmark research from the domains of spatial cognition and spatial computation. As demonstrated, these domains use a diverse set of variables when studying landmarks but it is questionable how useful these variables are within the domain of navigation support research. We review what role such technology is intended to play in users' everyday life and argue that in the most popular scenarios of use the dominant variables are not critical for evaluating navigation support systems. However, increasing the use of alternative – not yet crystallised – evaluation criteria is associated with methodological risks. One proposed way to mitigate them is an approach used in psychological science for pre-registering "postdiction" analyses. The paper aims to stimulate a debate on a corresponding consensus-based approach in our community.

2 Variables and evaluation criteria within the landmark literature

In a recently published literature review Yesiltepe et al. [27] aggregated 59 papers on the visibility of landmarks, selection of landmarks, location of landmarks, personal and emotional landmarks, gaze behaviour with relation to landmarks, and landmark saliency. We extend this literature review by classifying these papers into two main categories: (a) psychological research aimed at understanding how the mind works; and (b) computational research aimed at solving the problem of landmark identification based on human-specific criteria. This classification has been performed manually by the authors, based on the content of each paper. It is available in full at: `https://osf.io/ru8yb/`.

Next, we checked which dependent variables were reported in those papers' "Results" or "Evaluation" sections (when available) and classified those into higher-level groups of *evaluation criteria*. Table 1 presents evaluation criteria across the psychological and computational paper types.

Table 1 Evaluation criteria used across two types of papers treating on the problem of landmarks. Numbers in the brackets indicate how many papers of the given category applied the given criterion.

	Psychological papers	Computational papers
number of papers	39	18
evaluation criteria	spatial memory performance (21)	subjective salience judgement (5)
	navigational performance (12)	head/body alignment (1)
	properties of route descriptions (8)	navigational strategy (1)
	subjective salience judgement (5)	properties of route descriptions (1)
	gaze behaviour (distribution) (4)	
	quality of route descriptions (4)	
	self-localization performance (4)	
	gaze behaviour (fixations) (2)	
	variance of landmark placement (2)	
	head/body alignment (1)	
	navigational strategy (1)	
	stress (1)	

We identified 39 psychological papers, and 18 computational papers. Four papers remained unclassified: one literature review paper, one containing too little methodological detail, and two that focused on developing and evaluating a new navigation system but not on contributing to the understanding of psychological processes.

2.1 Psychological research

Psychological papers used 12 different types of evaluation criteria. The most popular one was spatial memory performance – this includes tests of spatial memory (survey, route, or landmark memory) where the reported variable is the accuracy or speed of response in a test. Spatial memory performance was studied in more than half of the reviewed papers. The second most-popular criterion was navigational performance – this includes measures of errors in, or optimality of, physical or virtual-reality-based wayfinding tasks. Navigational performance measurement was applied by almost 1/3 of all reviewed papers. The third most-popular evaluation criterion were properties of route descriptions – these are mainly analyses of sketchmaps produced in response to a task such as "describe the route to a stranger" (or similar), but *not* measuring the accuracy or quality of these descriptions. For instance, Peters [18] counted how many participants did vs did not include landmarks in their sketch map.

This analysis demonstrates that psychological papers seem to have converged on a set of measures: spatial memory performance and navigational performance are of interest to the highest number of papers.

2.2 Computational research

Computational papers used standardised evaluation criteria less often – out of 18 papers in this category we noted only 6 papers that applied any at all (one paper applied 3 of them). Of those 6 papers, 5 papers used participants' declared subjective judgement of landmark's salience (or landmark's "relevance") as their evaluation criterion. In the majority of cases, however, the most important evaluation criterion of the newly proposed computational solution is simply the fact that it works. Many papers demonstrated this by applying their method in a use case involving a real-world dataset.

This analysis demonstrates that research has investigated the landmark problem from diverse perspectives and with different motivations. Computational papers have diverse aims (e.g., mathematical–, engineering–, or data processing–oriented) and an empirical evaluation is not their main contribution. They maintain a connection to psychological literature but remain focused on own goals. Individual motivations of computational work might be useful, for example, to help develop methods for obtaining or processing previously underutilised datasets.

2.3 Navigation support research

However, there is another group of studies not well captured by this particular literature review: navigation support systems, also referred to as cognitive geoengineering applications [19]. These studies focus on supporting users in everyday task of navigation with the use of landmark-enhanced technology. Often, navigation support system research makes use solely of the same variables as employed by the strictly-psychological papers. For example, Wunderlich et al. [26] evaluated two different types of landmark-enhanced navigational instructions in comparison to standard no-landmark instructions. They measured spatial memory, navigational performance, and subjective mental load.

In some cases, navigation support systems research has considered different evaluation criteria as an add-on to spatial memory performance and navigational performance variables. For instance, Schwering et al. [20] designed a wayfinding-support system enhanced with on-screen and off-screen landmarks, with the aim of supporting user's orientation. In order to evaluate the approach they used measures of spatial memory performance. In addition, they discussed spatial memory variance – in particular the fact that that their wayfinding system seemed to have increased the consistency of pointing errors, regardless of their accuracy. Another example of such an approach is a paper by Smith et al. [21]. The authors implemented and tested navigation systems with off-screen landmarks inspired by those found in video games, and compared it to a standard Google Maps application in an in-situ experiment. In primary analyses they evaluated spatial memory performance through the accuracy and speed of pointing, as well as a map mark-up task. In addition, researchers analysed video recordings taken when participants performed their tasks in order to identify behavioural patterns such as: focusing on the environment vs on the phone, reorienting the device, frequency of glancing between the device and the environment, and signs of confusion. Yet another type of add-on variables used in evaluating navigation support research are measures of human-computer interaction. For instance, Li et al. [12] recorded the number of pans and zooms performed on the device during an in-situ navigational task.

This analysis demonstrates that spatial memory performance and navigational performance measures are prioritised and other measures are reported as secondary. A paper evaluating a new approach to landmark-enhanced navigation support is expected to report them, and ideally to show improvement in comparison to standard technological solutions. As our personal experience shows, when the dominant measures are not recorded in the experiment, or are reported with a downplayed importance, the paper raises concerns in the peer-review process. However, the goal of a new navigation support system technology may indeed be different from improving spatial memory or navigational performance of its user. Simply not decreasing spatial memory and navigational performance might be entirely satisfactory if the technology demonstrates other benefits or novelty.

In the publication process, however, defending such a null hypothesis is problematic. First, it requires deciding on one of few available (but conceptually different) Bayesian formulations of null-hypothesis testing [11]. Second, not highlighting measures accepted as standard in the field bears the tell-tale signs of potential "HARKing" (Hypothesizing After Results are Known) and "fishing expeditions" (performing many tests on variables of secondary importance until something significant is found) [1]. A paper that reports no significant improvement in spatial memory and navigational performance, but focuses on less popular variables (often measured with custom-built tasks), justifiably rises red flags of questionable research practices. One interpretation seems very likely: researchers had attempted to improve spatial memory or navigational performance with the proposed technology, failed to do so, and abused their secondary analyses long enough to finally find something significant that they now present as an achievement.

It is the position of this paper that the motivation standing behind developing navigation support systems is different from the categories described in the two previous sections, and that this type of research need not directly inherit evaluation criteria central to psychology and spatial computation. The open problems in the evaluation of navigation support systems are what the new criteria should be and how to give them due diligence in the process of investigation and peer review.

3 How to evaluate navigation support systems

We propose a vision of shifting the focus of navigation support research away from spatial memory and navigational performance; and instead moving towards standardising and promoting an alternative evaluation criterion. As such a criterion, we propose investigating whether the *cognitive task at hand has changed due to equipping the user with the new technology.* Below, we first justify why spatial memory and navigational performance should not be prioritised by new technology. Second, we justify why a change in the cognitive task is a relevant criterion for the scientific goal of developing novel technology. Third, we propose the means towards standardising and promoting it within the field inspired by the still scarcely used approach of pre-registering "postdiction" studies in the field of psychological science.

3.1 Counterarguments to prioritising spatial memory and navigational performance

Spatial memory performance and other criteria from strictly-psychological research are the de-facto seal of approval for navigation support systems. One line of argument behind this is a concern that spatial technologies lead to spatial infantilization [23, 14], i.e., to reducing our capabilities of performing given tasks without the technology. Thus, creating technology that increases (as opposed to diminishing) spatial knowledge seems to be a logical counterstrategy. Not only is it a direct evidence that a user learnt something new thanks to the technology; but it can also be an indirect evidence for intensified engagement with the environment, or deeper cognitive processing of spatial information.

A counterargument to this is that humans have always created tools to amplify the limitations of their own mind. These have resulted in some cognitive infantilization but also in societal benefits [5]. Ford et al. [5] describe the example of a calculator: we could argue that its introduction removed the need for learning how to count. But at the same time it allowed a vast group of untrained users to engage with activities requiring reliable counting; while those who want, need, or like to count mentally, can still do so. Evaluating the usefulness of a newly proposed navigation support systems with spatial memory or navigational performance is similar to evaluating the usefulness of a calculator by studying whether the user memorised the outcome, learnt how to perform similar calculations on their own, or how often they obtained the correct solution. The first two criteria are irrelevant – the presence of the calculator makes them obsolete. This is the reason to buy one. The latter is a given – we expect any calculator to provide the correct response at all times. Potential problems might arise due to human-computer interaction (e.g. errors in input) – those would be relevant and can be fixed but pertain to the interface design, not to the role or usefulness of this technology to the society. If the goal of the calculator is to empower its users, and not to teach them (we create other tools for that), then variables related to learning are of secondary importance.

Another popular line of argument for evaluating spatial technologies with strictly-psychological variables is the cognitive geoengineering argument that the spatial concepts used by the systems must *match* spatial concepts used by its user [19]. As Table 1 demonstrates, finding a *match* between the algorithm and human preference is the most popular evaluation criterion for strictly-technological research. This is certainly valuable, as it demonstrates a success of interdisciplinary spatial cognition community: It is routine for computer scientists to understand, target, and evaluate human-centred issues.

But it is worth considering how close a *match* we aspire to? Let us differentiate between two levels of such a *match* in the domain of automated landmark identification: a *soft* and a *hard match*. We would consider a *soft match* all instances when the human understands the system referring to some object as a "landmark" – in a similar way we understand when a stranger (with different personal background, spatial knowledge, and preferences) refers to something that they consider a "landmark". We would consider a *hard match* such instances of automatically identified landmarks that would be identical to what the user themselves would identify (in the given context, emotional state, and within the given task).

A soft match is certainty sufficient for sustaining everyday communication. Simply referring to some perceivable object in navigational instructions makes it a landmark [15], even if the object does not exist, like the famous "etak" islands [8]. Thus, the goal of landmark-enhanced navigation support need not be to mimic the human mind by detecting the same landmarks as the user would. In fact, given well-defined criteria and a large database, a computer algorithm will outperform humans in consistently detecting landmarks or classifying their salience (within the definition formalised by the researcher). When no landmark is available in the immediate environment, the system has the means to compensate for this, e.g., by visualising information that is not directly perceivable in the environment, or by imposing schematised order and salience onto a chaotic or uncharacteristic surrounding [6]. This potential of landmark-detection algorithms seems to be underutilised in designing landmark-enhanced navigation support systems. Instead, researchers focused on the challenge of "biasing" those consistent results in order to match human biases, for instance on the base of emotional judgements in landmark preference.

3.2 Evaluating the human-computer cognitive system

It is the position of this paper that the goal of navigation support systems research is not to bridge the gap between solutions of the landmark computation performed by the algorithms and the ways in which humans do it. This may be useful for other research goals: a system achieving a "hard" match between digital data and human cognition would be a theoretical achievement and may have some practical use. But within navigation support systems, constructing technology mimicking how the mind works might be as difficult, as it is unnecessary. Instead, we propose to evaluate the cognitive system consisting of the technology and its user, in line with the distributed cognition framework of cognitive engineering [4]. Within this theoretical approach, humans and the tools they use are viewed as a separate kind of a cognitive system: one that has different properties than any of its parts. Such a newly established cognitive system should not be evaluated based on criteria specific to its sub-parts. A human equipped with a navigation support system does not need to identify or remember landmarks because the system can do it for them. So what should we evaluate instead?

A first step could be deciding on (and declaring) the intended role of the landmark-enhanced technology in human everyday life. Hermann [7] presented a framework of cognitive technology, classifying technological applications based on their role in supporting or substituting human cognitive processes. Eight types of technologies are listed, for instance a "cognitive prosthesis". The role of such technology is to externalise a cognitive function, similarly to how a reminder in a smartphone externalises the need for using memory [7]. This category in Hermann's framework was further subdivided by others [2] to highlight an important difference between a "cognitive prosthesis" and a "cognitive amplifier".

The difference is that the role of a prosthesis is to bridge the gap between some characteristic or ability of an individual and an accepted average, or a standard previously achievable by experts. For instance, a pair glasses "brings" the eye-sight back to an agreed standard [2].

This could be one role of landmark-based systems: a subset of the population that cannot easily navigate with other existing tools could make good use of a "cognitive prosthesis". One such group might be visually impaired [17].

Compared to a cognitive prosthesis, a cognitive amplifier acts by allowing people to complete more advanced tasks, regardless of the agreed standard. Importantly, this happens not because tools increase computational powers of the human mind, but rather because they change the cognitive task that is being performed [2, 4]. One strategy for changing the nature of a cognitive task is to delegate what used to be an expensive mental calculation onto a simpler perceptual ("pre-cognitive") system [2]. The success of the turn-by-turn navigation systems stems from the fact that they substituted complicated cognitive operations (e.g., planning and remembering actions based on an allocentric perspective of a map) with simple perceptual ones: "take action when you see or hear a notification". This is not detectable by measuring variables specific to the user-part or the computer-part of the joint human-computer system.

In fact, human cognitive performance in spatial learning could decrease because it is being delegated to the technology. Navigational performance may lose relevance, because the human-computer system can achieve satisfactory performance in a wast majority of cases. These measures are not indicative of the entire human-computer systems because the distribution of cognitive tasks between the human and the computer changes in various human-computer systems. This phenomenon is schematically depicted in Figure 2. If to focus on performance measures of the entire human-computer system, a human with no computer (Fig. 2a) has potentially the lowest navigational efficiency but the highest share of cognitive tasks. A human with a map (Fig. 2b) can rely on the technology instead of own spatial memory, thus the navigational efficiency of the total system is higher, but the share of the human in the cognitive work is lower within the system. What is not captured clearly by any standard measure is that the *cognitive task* has changed: the user can perform new cognitive tasks (e.g., plan a route on a symbolic representation of the map) and still reach the destination. A human with a turn-by-turn wayfinding system has a higher navigational efficiency, but the share of the human-part in the computational work of the system is even lower compared to a map. The cognitive task of the human has changed because one does not need to plan turning actions (Fig. 2c). A system detecting and communicating personalised, context-specific landmarks instantly understandable by the user (Fig. 2d) is likely to reduce the human's demand for cognition because one will be able to perform the task with even less effort spent on identifying turning points. Yet, it is yet unclear whether this will change the *cognitive task* performed by the user. The potential usefulness and acceptance of this technology is also unclear. Future technological applications (Fig. 2e) are unlikely to improve the navigational efficiency any further. The involvement of the human in the cognitive work of the system might change depending on the purpose of the technology, but should not be the key variable based on which we evaluate its potential, novelty, or success.

What is needed is a way of evaluating the success (or novelty, or potential) of newly proposed navigation support systems that would depart from strictly-psychological or strictly-technological variables such as those documented in Table 1. We suggest two criteria: (1) considering the role of the technology in the society (e.g., within Hermann's framework) with one potential aim being to adapt some technological solution to a new societal role; and (2) documenting a change in the type of the cognitive task required from the human-part of the human-computer system; this would make it possible to draw qualitative distinctions between two systems where the human involvement in computation is similar or decreases.

With regard to (1), consider the societal role that a landmark-enhanced navigation support system might have inside autonomous vehicles. Although not required for navigation, it could still serve to provide a feeling of comfort or security during a trip through an unknown area, in which case it would not be a "cognitive amplifier", but a "cognitive self-care facilitator" [7]. Another example are "cognitive trainers" that are purposefully designed to train cognitive skills. It would be relevant to evaluate the performance of such system by measuring gains in spatial memory. Note, however, that the goal of most navigation support systems for everyday use is not to be "cognitive trainers".

With regard to (2), this is where taxonomies and task classifications might be particularly helpful; such as the taxonomy of wayfinding tasks proposed by [25] and extended to technology-supported wayfinding by [20]. One explicit goal of novel technology could be to change the distribution of cognitive tasks between the user and the system, i.e., to allocate the user and the newly proposed system in a different taxon. However, it is unclear how to assess whether two systems are enabling, or encouraging the user to perform different cognitive tasks. In the next section we propose a step towards building up an approach for this kind of evaluation.

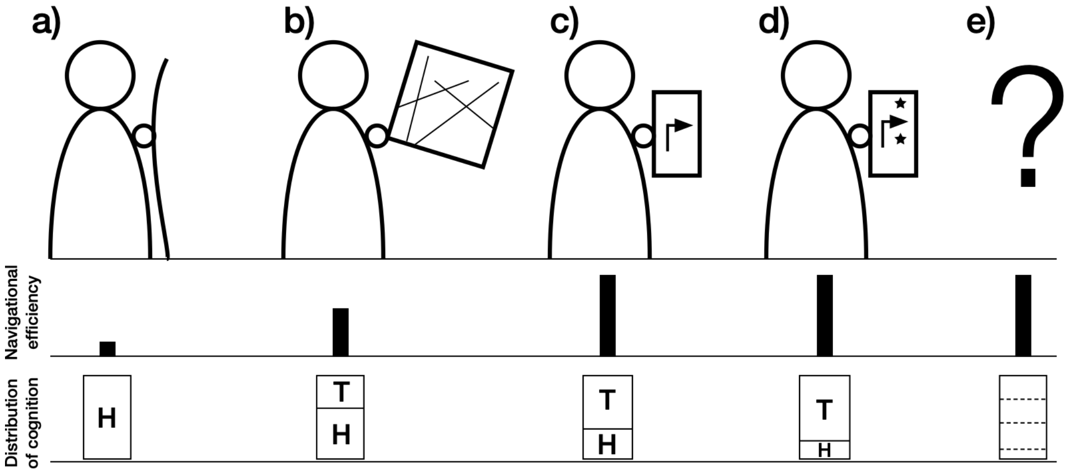

Figure 2 A schematic depiction of various human-computer cognitive systems in wayfinding with relation to standard evaluation criteria and the distribution of cognitive tasks between the human (H) and the technology (T). (a) A human with a walking stick and no computer. (b) A human with a map. (c) A human with a turn-by-turn wayfinding system. (d) A human with a system detecting and communicating personalised, context-specific landmarks instantly understandable by the user. (e) Future technological applications are unlikely to improve the navigational efficiency any further. How should they change the cognitive task performed by the user?

3.3 Pre-registration as a way to standardise and promote new criteria

Understanding how to measure the two above-mentioned research goals is an open challenge. We suggest to approach it from the perspective of exploratory research. This is different from the approach typically applied by papers introducing a new navigation support technology. The default approach is to treat it as confirmatory research, where researchers use theories of spatial cognition (e.g., a generalised assumption that salient landmarks are easier to memorise), develop some technology utilising this theory (e.g., a system that shows salient landmarks), and design an experiment to confirm if the assumption is true (e.g., if participants using the system had statistically significantly better memory performance with regard to the newly visualised salient landmarks).

This approach has two impediments. First, it binds together the test of the theory with the evaluation of the technology: if a said system would not improve participants' memory, it is impossible to conclude whether this is due to the theory being wrong (i.e., in this particular context salient landmarks are not easier to remember), or whether the technology is not a good application of this particular theoretical insight (e.g., something about the visualisation of landmarks in this particular application hinders the memorisation effect). This would not be a problem if theories we use were strongly verified in practical context of use but this is not the case for most theories in behavioural and psychological science [9]. Second, this approach creates an "all-or-nothing" narrative about the success of the proposed technology: if the new technology does not improve the memorisation of landmarks, the approach has failed, and other measures are reported at most as an interesting side-effect in secondary analyses. Reviewers' and readers' scepticism towards this kind of secondary analyses in confirmatory studies is justified. From the statistical viewpoint, one dataset cannot be used to both generate and validate a hypothesis because a dataset used to *find* a phenomenon is not going to falsify it [24].

An alternative approach is to explicitly treat validation of new navigation support systems tested with participants in naturalistic scenarios as exploratory research. The goal of exploratory research is not to validate a hypothesis but to describe interesting properties of the data, determine which (tentative) findings are interesting, and to propose steps forward for future research [24]. This means that what we are used to call an "evaluation" of new technological propositions, would in fact be an exploratory data analysis project. From the standpoint of cumulative science, a declared exploratory study has a higher value than an undisclosed exploratory practice looking like a confirmatory test [16, 24].

This is a particularly challenging issue for our discipline. Creating a novel navigation support system and testing it in a naturalistic scenario is proceeded by decades of research in psychology and spatial computation; it is based on theories confirmed in numerous experiments; it utilises advanced computational methods that must have matured before they became usable in practical contexts. To call such research "exploratory" does not give it justice. Yet, is seems methodologically more appropriate to treat it as exploratory than as confirmatory analysis. For this reason, for evaluating navigation support systems we propose to adopt the term "postdiction" – defined as research undertaken after the data is known with the goal of explaining why it occurred (also referred to as "research in the context of discovery" [16]).

Dirnagl [3] compares doing such exploratory studies to the endeavour undertaken by sailing expeditions in the times when maps were still incomplete. These attempts were pre-registered, in the sense that the goal had to be specified. Yet, the retrospective value of the expeditions is not only whether they did or did not reach the goal. Partial discoveries of predecessors were valid – even if only fragmentary – reasons to pursue or avoid particular directions in the future. But what made the cumulative learning from this expeditions possible were logs kept from the journeys. Pre-registering research as exploratory, publishing study protocols, and keeping a public log of exploratory analyses is the modern science equivalent of such expeditions.

For landmark-enhanced navigation support systems we suggest the following ways for enhancing the cumulative effort of discovering novel and useful technology:

1. A declaration of the intended role of the newly proposed technology (e.g., a cognitive amplifier or a cognitive trainer, within the above described framework [7]).
2. A pre-registration of the claim that the newly proposed system *will not significantly decrease navigational performance*. Since navigation support is already sufficiently supporting performance, expecting that new systems will further improve on this criterion

seems unjustified. In some cases, a similar claim might be made for spatial memory measures – not all contexts require that spatial memory of the travelled route increases. Pre-registering this claim makes it explicit, open, and traceable, that researchers were not focused on increasing measures accepted as dominant in the field. This can assist authors and readers in shifting attention to other aspects tested in the study.

3. A pre-registration of intended postdiction analyses aimed at detecting whether the cognitive task performed by the human in the joint human-computer navigation system has been changed by the newly tested technology. Such analyses are sensitive to the specificity of the dataset, therefore the highest value lies in detecting a change by using an approach developed by other researchers on another dataset. These are not unknown in the domain of landmark research but a set of standardised methods that would be appropriate for this research goal has yet to crystallise. Arguments have been made for *detecting strategies and variance* [10] in navigational behaviour and problem-solving, instead of methods for quantifying performance.

4 Conclusion

Landmark research spans decades, but the use of landmarks in everyday navigation support technology to date has neither been realised in common applications, nor proven necessary. Hermann [7] argued that demonstrating a successful application can be an ultimate check for basic research – a check that landmark-enhanced navigation support systems have not yet passed. The claim presented in this paper is that the reason is not a technological limitation but rather the way in which navigation support research is evaluated – with an over-reliance on evaluation criteria specific to strictly-psychological research.

Two suggestions have been made for moving beyond the impasse: (1) re-considering the role of landmark-enhanced navigation support systems in everyday life of their users; and (2) standardising the evaluation of a phenomenon difficult to capture with traditional measures – whether the technology changed the cognitive task performed by its user. As no clear set of methods for detecting the latter are agreed upon, we suggest *pre-registering exploratory analyses* prior to collecting the data as a means of highlighting this goal. Methods for detecting change and variance in cognitive or behavioural strategies, developed by researchers on other datasets, should be prioritised.

This turn to exploratory analyses is not an attempt to diminish or criticise the quality of navigation support research but an attempt to set it apart from priorities central to strictly-psychological or strictly-computational research. Moving towards this goal requires the development of evaluation criteria better reflecting the societal potential (or scientific novelty) of the new technology. In the spirit of "use-inspired basic research" [22] we are strongly convinced about the potential that such technology can also feed back into basic science; for instance by shifting research focus onto cognitive processes and strategies that gain importance in everyday life inevitably assisted by technology.

References

1 Chittaranjan Andrade. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry*, 82(1), 2021. doi:10.4088/jcp.20f13804.

2 Richard Arias-Hernandez, Tera M Green, and Brian Fisher. From Cognitive Amplifiers to Cognitive Prostheses: Understandings of the Material Basis of Cognition in Visual Analytics. *Interdisciplinary Science Reviews*, 37(1):4–18, 2013. doi:10.1179/0308018812z.0000000001.

3 Ulrich Dirnagl. Preregistration of exploratory research: Learning from the golden age of discovery. *PLOS Biology*, 18(3):e3000690, 2020. doi:10.1371/journal.pbio.3000690.

4 Itiel E Dror and Stevan Harnad. Offloading cognition onto cognitive technology. In *Cognition Distributed: How Cognitive Technology Extends Our Minds*, volume 16, pages 1–23. John Benjamins Publishing, 2008. arXiv:0808.3569.

5 Kenneth M. Ford, Clark Glymour, and Patrick J. Hayes. On the Other Hand ... Cognitive Prostheses. *AI Magazine*, 18(3):104, 1997. doi:10.1609/aimag.v18i3.1317.

6 Marcelo De Lima Galvão, Jakub Krukar, Martin Nöllenburg, and Angela Schwering. Route schematization with landmarks. *Journal of Spatial Information Science*, 21:99–136, December 2020. doi:10.5311/JOSIS.2020.21.589.

7 Douglas J. Hermann. The potential of cognitive technology. In W. Richard Walker and Douglas J. Herrmann, editors, *Cognitive Technology: Essays on the Transformation of Thought and Society*, pages 5–19. McFarland & Co, Jefferson, N.C, 2005.

8 Edwin Hutchins. *Cognition in the Wild*. The MIT Press, Cambridge, MA, US, 1995.

9 Hans IJzerman, Neil A. Lewis, Andrew K. Przybylski, Netta Weinstein, Lisa DeBruine, Stuart J. Ritchie, Simine Vazire, Patrick S. Forscher, Richard D. Morey, James D. Ivory, and Farid Anvari. Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11):1092–1094, 2020. doi:10.1038/s41562-020-00990-w.

10 Alan Kingstone, Daniel Smilek, and John D. Eastwood. Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3):317–340, 2008. doi:10.1348/000712607x251243.

11 John K. Kruschke. Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, 6(3):299–312, 2011. doi:10.1177/1745691611406925.

12 Rui Li, Amichi Korda, Maurin Radtke, and Angela Schwering. Visualising distant off-screen landmarks on mobile devices to support spatial orientation. *Journal of Location Based Services*, 8(3):166–178, 2014. doi:10.1080/17489725.2014.978825.

13 K Lynch. *The Image of the City*. MIT Press, 1960.

14 Daniel R Montello. Cognitive Research in GIScience: Recent Achievements and Future Prospects. *Geography Compass*, 3(5):1824–1840, 2009. doi:10.1111/j.1749-8198.2009.00273.x.

15 Daniel R. Montello. Landmarks are Exaggerated. *KI - Künstliche Intelligenz*, 31(2):193–197, 2017. doi:10.1007/s13218-016-0473-5.

16 Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):201708274, 2018. doi:10.1073/pnas.1708274114.

17 Rajchandar Padmanaban and Jakub Krukar. Increasing the density of local landmarks in wayfinding instructions for the visually impaired. In Georg Gartner and Haosheng Huang, editors, *Progress in Location-Based Services 2016*, pages 131–150, Cham, 2017. Springer International Publishing.

18 Denise Peters, Yunhui Wu, and Stephan Winter. Testing Landmark Identification Theories in Virtual Environments. In Christoph Hölscher, Thomas F. Shipley, Marta Olivetti Belardinelli, John A. Bateman, and Nora S. Newcombe, editors, *Spatial Cognition VII*, Lecture Notes in Computer Science, pages 54–69, Berlin, Heidelberg, 2010. Springer. doi:10.1007/978-3-642-14749-4_8.

19 Martin Raubal. Cognitive engineering for geographic information science. *Geography Compass*, 3(3):1087–1104, 2009. doi:10.1111/j.1749-8198.2009.00224.x.

20 Angela Schwering, Jakub Krukar, Rui Li, Vanessa Joy Anacta, and Stefan Fuest. Wayfinding Through Orientation. *Spatial Cognition & Computation*, 17(4):273–303, 2017. doi:10.1080/13875868.2017.1322597.

21 Alastair D Smith, Gary Priestnall, and Juliette Cross. Supporting spatial orientation during route following through dynamic maps with off-screen landmark persistence. *Spatial Cognition & Computation*, pages 1–28, 2021. doi:10.1080/13875868.2021.1985122.

22 Donald E Stokes. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press, 1997.

23 Tyler Thrash, Sara Lanini-Maggi, Sara I. Fabrikant, Sven Bertel, Annina Brügger, Sascha Credé, Cao Tri Do, Georg Gartner, Haosheng Huang, Stefan Münzer, and Kai-Florian Richter. The Future of Geographic Information Displays from GIScience, Cartographic, and Cognitive Science Perspectives (Vision Paper). In Sabine Timpf, Christoph Schlieder, Markus Kattenbeck, Bernd Ludwig, and Kathleen Stewart, editors, *14th International Conference on Spatial Information Theory (COSIT 2019)*, volume 142 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:11, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.COSIT.2019.19.

24 Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit. An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6):632–638, 2012. doi:10.1177/1745691612463078.

25 Jan M. Wiener, Simon J. Büchner, and Christoph Hölscher. Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, 9(2):152–165, 2009. doi:10.1080/13875860902906496.

26 Anna Wunderlich, Sabine Grieger, and Klaus Gramann. Landmark information included in turn-by-turn instructions induce incidental acquisition of lasting route knowledge. *Spatial Cognition & Computation*, pages 1–26, 2022. doi:10.1080/13875868.2021.2022681.

27 Demet Yesiltepe, Ruth Conroy Dalton, and Ayse Ozbil Torun. Landmarks in wayfinding: a review of the existing literature. *Cognitive Processing*, 22(3):369–410, 2021. doi:10.1007/s10339-021-01012-x.