Augmented Reality as a Visualization Technique for Scholarly Publications in Astronomy: An Empirical Evaluation

Jane L. Adams*
Northeastern
University

Laura South*
Northeastern
University

Arzu Çöltekin†
University of Applied Sciences
and Arts, Northwestern
Switzerland (FHNW)

Alyssa A. Goodman‡
Harvard
University

Michelle A. Borkin*
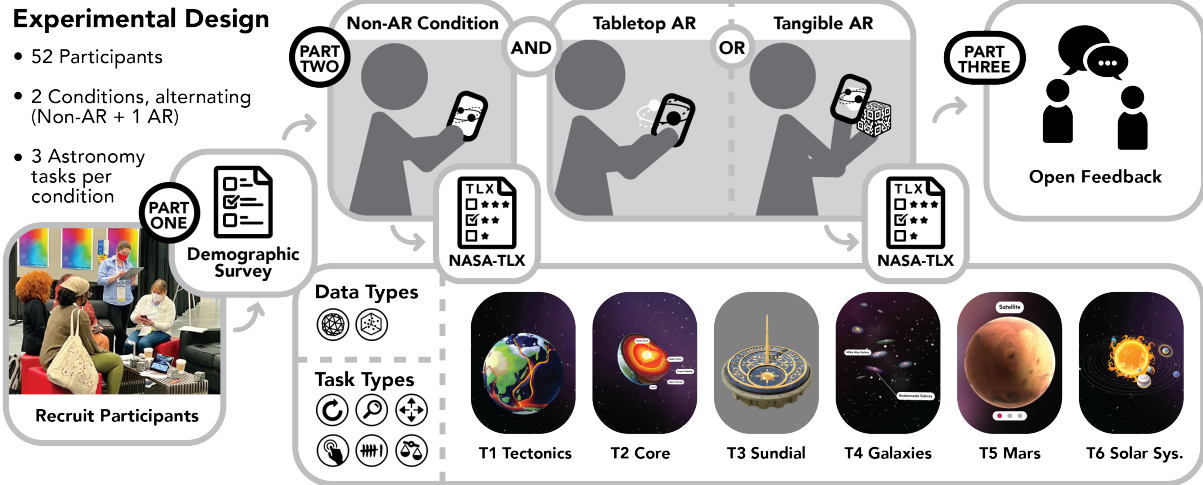Northeastern
University

Figure 1: Experimental design: **Part 1** A survey on participants' expertise, experience with AR, and demographic information, **Part 2** Two sets of three tasks in each the non-AR and one of the two AR conditions, along with NASA-TLX workload questionnaires, and **Part 3** Open feedback from participants.

ABSTRACT

We present a mixed methods user study evaluating augmented reality (AR) as a visualization technique for use in astronomy journal publications. This work is motivated by the highly spatial nature of scientific visualizations employed in astronomy, including spatial reasoning tasks for hypothesis generation and scientific communications. In this 52-person user study, we evaluate two AR approaches (one traditional tabletop projection and the other with a 'tangible' aid) as spatial 3D visualization techniques, as compared to a baseline 3D rendering on a phone. We identify a significant difference in mental and physical workload between the two AR conditions in men and women. Qualitatively, through thematic coding of interviews, we identify notable observed differences ranging from device-specific physical challenges, to subdomain-specific utility within astronomy. The confluence of quantitative and qualitative results suggest a tension between workload and engagement when comparing non-AR and AR technologies. We summarize these findings and contribute them for reference in data visualization research furthering novel scientific communications in astronomy journal publications.

Index Terms: Human-centered computing—Visualization— Empirical studies in visualization——Human Computer Interaction (HCI)—Interaction paradigms—Mixed / augmented reality

*e-mail: {adams.jan, south.l, m.borkin}@northeastern.edu
†e-mail: arzu.coltekin@fhnw.ch
‡e-mail: agoodman@cfa.harvard.edu

1 INTRODUCTION

The fields of astronomy and astrophysics require analysis of large amounts of spatial data, such as star positioning [25], nebulae structure [27], or simulations of dark matter [22]. Consequently, researchers in these fields often use 3D visualization techniques such as 3D scatterplots, isosurfaces, and volume renderings to wrangle, explore, and understand astronomical data [24]. These visualizations are most commonly created using non-immersive 2D laptops or desktop monitors [2, 18]. Despite recent advances in 3D visualization techniques such as augmented reality (AR), researchers are generally restricted to representing complex spatial data using conventional static figures suitable for publication in print journal papers. Such images often consist of static screenshots of the 3D representation for a specific angle (or a small multiples with multiple views), contour plots, or exclusion entirely of the 3D visuals. Some alternative approaches which preserve the 3D representation as well as provide interactivity include links to 3D visualizations hosted on the web, or embedded 3D figures directly in PDFs [3, 16]. However, these alternative approaches require technical skills in order to author the figures, and web-based visuals typically do not enable a reader to see the 3D visualization effectively in the context of the journal paper itself (i.e., need to navigate or look away from the paper). The considerations above leaves us with the question *How can we more effectively include interactive 3D visualizations as journal paper figures?*

A motivating example of a potentially effective interactive 3D journal figure was published by Bialy et al. (2021), in which they present a series of conventional 2D screenshots of 3D visualization along with QR codes that, when followed, take the reader to Augmented Reality (AR) visualizations. These interactive 3D visualizations potentially enable the reader to more effectively understand and see their discovery of the Per-Tau "Supershell" [5]. The state of the art workflow right now for journal publication of AR figures is for users

to either author in Plotly https://plotly.com, or export from visualization software to Plotly (e.g. from Glue http://glueviz.org) [21], and then import from Plotly into CoSpaces [15] (https://cospaces.io/). The media has lauded these interactive figures, several of which have been embedded in high-profile publications [32, 34, 41]. Discoveries such as the one visible in Bialy et al.'s interactive 3D visualization are not immediately evident from 2D screenshots alone [5]. Anecdotally, this new AR journal paper figure for 3D visualization was a success based on feedback from the authors and paper readers. However, in a recent study, Ambühl & Fluri et al. (2022) implemented an AR solution in the context of astronomy, and user-tested the AR solution for analytical tasks with 16 participants, reporting that for such tasks AR was *not* effective (i.e., a desktop 3D version was more efficient and effective), but all participants liked AR [1]. Preliminary results from this study also suggest that spatial abilities might play a role, and authors conclude that AR might have value in experience-based insights rather than analytical tasks. It is however, still to be established: *Is an interactive AR figure really better than a more conventional 3D visualization on a screen? What is intuitive or challenging for the user?*

To answer these questions we conducted a mixed-methods user study to assess the workload and feasibility of three technologies to display interactive 3D visualizations of astronomical data: conventional 3D rendering on a digital screen (**Non-AR**), "classic" augmented reality with 3D visualizations placed into the user's environment (**AR Tabletop**), and AR projected onto a holdable "Merge Cube" (https://mergeedu.com/cube) tangible device (**AR Tangible**), as shown in Fig. 1. We conducted the user study at the 241st Meeting of the American Astronomical Society with 52 prospective users. The primary objective of our study is to evaluate the strengths and weaknesses of each technology, providing important groundwork for future research on how to best display and interact with 3D visualizations, and offer evidence of feasibility of AR enhanced journal paper figures. Taken together, our findings contribute to data visualization and XR knowledge via an empirical evaluation which examines the feasibility and usability of three different methods for displaying and interacting with 3D visualizations. We present findings related to perceived difficulty, demographic differences, and qualitative feedback which can inform subsequent studies and software development at this intersection of astronomy, data visualization, and extended reality.

2 RELATED WORK

AR is a technology that superimposes digital information onto the real world, allowing users to interact with and visualize data in an immersive and contextualized manner. AR has been used in a variety of spatial contexts [10], e.g., situated analytics [8,14], same-scale contextual geometry [40], and geography [19]. In the context of AR, a tangible, such as the "Merge Cube", refers to a physical object or artifact that is recognized and tracked by AR technology in real-time. Tangibles are typically designed to serve as interactive interfaces for users to manipulate and control virtual objects or information displayed on a digital device. Tangibles can take many forms, such as physical markers, tags, cards, or other objects that are uniquely identifiable and detectable by AR software [6]. Tangible AR has been used to visualize and manipulate 3D data in educational games [13], Embodied Axes [11], chemistry journal publications [39], and courseware [9], among others. Prior lab studies have compared non-tangible AR against desktop conditions [2] and tested the usability of different tangibles [30], demonstrating that there are some issues with AR. Specifically, readability differs based on encoding; and well-known concerns related to 3D visualization, such as occlusion, are also present in AR [40]. It has been shown that AR might continue to be outpaced in user ease relative to non-AR conditions [1, 2]. However, none of the previous work address specifically a rigorous comparison between conventional AR, AR with a tangible, and 3D on a phone with expert participants.

The user study presented in this paper goes beyond the above-mentioned prior work by evaluating across three technologies (Tangible AR, Tabletop AR, and non-AR as a control) with a large (n=52) participant pool, using only a cell phone and the Merge Cube tangible, a lightweight, portable, and affordable solution. Although there is some prior work demonstrating the feasibility of the Merge Cube tangible in education for teaching science [31, 35, 36] and medicine [37] no formal quantitative study has previously evaluated the Merge Cube besides the preliminary work by Ambühl & Fluri et al. [1] that compared Merge Cube to a desktop 3D solution. Finally, no prior work has specifically examined the use of AR (with or without tangibles) for visualization in the context of astronomy journal publications.

3 EXPERIMENTAL DESIGN

Protocol To evaluate and explore the user experience with different technologies for displaying astronomical 3D visualizations as journal figures, we conducted a mixed-methods study incorporating both quantitative and qualitative measures. The mixed (2x2 within-between subjects) design enabled us to compare the non-AR versus AR conditions (within-subject) as well as the workload of the two AR technologies (between-subject). The study had three parts: **Part 1** a survey to gather information about the participant's expertise, experience with AR, and demographic information, **Part 2** structured questions with tasks for the participant to complete with assigned technologies to measure workload, and **Part 3** semi-structured interview questions and post-study feedback. Prior to commencing the experiment, participants were briefed and asked to sign a consent form. The study took ∼20 minutes to complete, and participants were compensated after the study with a Merge Cube (retail value of $25). In the Supplemental Materials at https://osf.io/xazr3/, we provide the study protocol (including questions and stimuli), results, and analysis code. Our study preregistration is at https://osf.io/eukyp. This study was approved and executed in accordance with Northeastern University's Institutional Review Board.

Participants For better ecological validity, we chose to recruit participants from the astronomy and astrophysics research community (i.e., prospective authors and readers of 3D figures in astronomy journals). The study was conducted in the exhibit hall of the 241st Meeting of the American Astronomical Society (AAS) in Seattle, WA on January 8-12, 2023. Our 52 participants consisted of 25 men, 24 women, and 3 non-binary or gender-fluid participants, with a diverse range of expertise and ages. Full information about participant demographics are included in the Supplemental Materials.

Tasks Using a Latin squares design we split the participants evenly into (1) an AR condition (either Tabletop or Tangible), (2) the condition order (i.e., whether they see non-AR or AR first), and (3) the task ordering. The study consisted of six tasks evenly spread across three task categories: *panning*, *rotation*, and *zoom*. The tasks were rooted in the fundamental interactions associated with both how a user navigates a 3D visualization [4], as well as the data analysis tasks most common for astronomy visualizations including selection and comparison [24]. All participants used the same mobile phone and sat in the same chair to keep the environmental conditions constant.

Stimuli Each task in the study was associated with a specific dataset. The **data**, and its 3D rendering, were provided in and used with the mobile app Merge Explorer [28]. As Bialy et al. [5] is the only astronomy journal paper published with a 3D AR figure at the time of writing, we used astronomical 3D visualizations from the Merge Explorer app as experimental stimuli. Although the visuals were designed for K-12 education, the underlying data is based on real astronomy data. Also, to prevent confusion, as part of the study protocol, participants were informed that these visualizations were designed originally for education but are representative of the types of visuals and data that are found in journal publications. The datasets were chosen specifically to be conducive to the study's tasks as well as encompass both isosurface visualizations and 3D scatter plots [24], as shown in Table 1. Finally, as we anticipated a diversity of expertise in astronomy subfields, we chose stimuli easily understood by any astronomer, and comprised of clear simple tasks, as shown in Fig. 1. The specific task questions are provided in the study protocol in Supplemental Materials, along with video demonstrations of each task.

Evaluation For the main (Part 2) portion of our study, to quantitatively measure task workload we utilized the NASA-TLX workload scale [17]. The scale's six subscales (mental demand, physical

Table 1: A table showing the different stimuli and their associated data type and task types.

Task	Data Type	Task Type(s)
T1	Isosurface	Rotation, panning, counting
T2	Isosurface	Rotation, counting
T3	Isosurface	Rotation
T4	3D Scatter Plot	Rotation, panning, counting
T5	Isosurface	Zooming, panning, comparison
T6	3D Scatter Plot	Rotation, zooming, selection

demand, temporal demand, performance, effort, and frustration) are weighted and combined to yield an overall workload score [17]. For the qualitative analysis of the study (i.e., observations during Part 2 and verbal feedback in Part 3), we used grounded theory with thematic analysis and axial coding [33].

Analysis We had two hypotheses for the quantitative portion of the study: **Hypothesis 1 (H1): AR will require a lower workload than Non-AR technology.** We hypothesized that both AR conditions would be lower workload than the non-AR condition because it would be more intuitive to interact with the 3D visualization compared to a flat display. **Hypothesis 2 (H2): Tabletop AR and Tangible AR will perform equally well in terms of workload.** We hypothesized the workload ratings of both AR conditions to be approximately equal in workload due to their visual and interactive similarities. Additionally, based on prior research about differences in performance with certain spatial tasks across genders (e.g., [26]), we anticipated we would observe a significant difference in workload scoring between conditions controlled for **gender** for both H1 and H2.
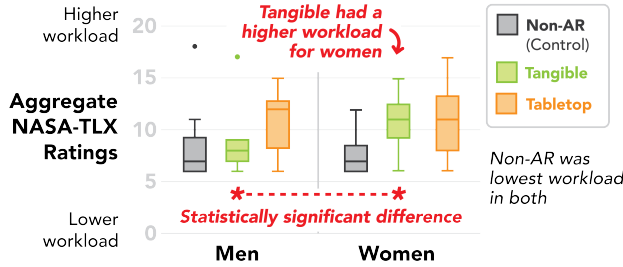

Figure 2: Quantitative NASA-TLX aggregate measures show that the non-AR condition was the lowest workload. The tabletop AR condition trended towards highest workload on average. Women experienced greater workload with the tangible AR device than the non-AR device.

4 RESULTS

Hypothesis 1 (H1) Non-AR had a lower workload score than both AR technologies for women, whereas there was no significant difference for men. We conducted paired-sample t tests to compare the self-reported TLX scores for control (i.e., non-AR 3D on smartphone) and AR (i.e., tangible and tabletop) technologies. To control for the effect of gender on participant responses, we conducted separate two-sided paired-sample t tests for men and women. We had three non-binary or gender fluid participants, which were excluded from the gender-based statistical analysis, but included in qualitative analysis. The assumption of normality in the distribution of paired differences was verified for men and women using the Shapiro-Wilk test ($p = 0.055$ and $p = 0.08$, respectively) and by visual inspection of the results. To account for the issue of multiple comparisons introduced by examining men and women separately, we set $\alpha = 0.025$ as our threshold for significance. As shown in Fig. 2, the non-AR system (i.e., 3D on phone) required a significantly *lower workload* than the AR systems among women ($M = -3.29, t(23) = -5.00, p < 0.0001$). No significant difference was detected among men between the non-AR and AR technologies for self-reported TLX scores ($t(24) = -1.81$, $p = 0.0815$). The average paired difference in TLX scores among men ($M = -1.44$) was in favor of the non-AR systems, following the same

pattern that was observed with women, although this effect was not found to be statistically significant. These results are shown in Fig. 2.

In our qualitative analysis, we found that tasks in the non-AR condition were solved more quickly (though not timed). Participants in this condition frequently returned the phone to the tester upon task completion, potentially indicating reduced engagement. Conversely, in both AR conditions, users often continued to explore the visualization until prompted by the tester to begin the next task, suggesting heightened engagement. Usability challenges varied across all three technologies. Zooming too fast (2-finger pinch) in the non-AR condition, especially in T5 Mars, was a common problem, causing users to collide with the planet, leading to disorientation. Additionally, participants faced difficulty closing dialogue boxes in all conditions, possibly due to similar navigation and selection motions.

Hypothesis 2 (H2) There was no significant difference between Tabletop AR and Tangible AR, though Tabletop averaged a higher workload score. We intended to conduct an independent samples t test to compare TLX scores for the between-subjects factor of *which* AR technology participants used. The assumption of normality was successfully verified with the Shapiro-Wilk test and by visual inspection for both Tabletop AR and Tangible AR among women ($p = 0.572$ and $p = 0.202$, respectively). The assumption of homoscedasticity was verified using the Levene test for both men and women ($p = 0.441$ and $p = 0.279$). Responses from men in the Tabletop AR group were approximately normally distributed ($p = 0.751$), but the mens' Tangible AR group failed to pass the Shapiro-Wilk test for normality ($p = 0.0003$). Therefore, the nonparametric Mann-Whitney test was used to compare Tabletop AR and Tangible AR among men. The parametric independent samples t test was used to compare women groups. Descriptive statistics suggest that men found the Tangible AR to be lower workload on average than the Tabletop AR ($\bar{x}_{Tangible,M} = 8.54$, $\bar{x}_{Tabletop,M} = 10.80$), as shown in Fig. 2, however we fail to reject the null hypothesis that the Tangible AR and Tabletop AR samples had the same distribution based on the results of the Mann-Whitney test ($U = 35.0, p = 0.063$). There is insufficient evidence to conclude that men experienced lower workload with either of the AR technologies. Similarly, women did not report significantly different TLX scores between the two AR systems ($t(20) = 0.06, p = 0.950$).

Qualitative Qualitatively, each AR technology presented unique usability challenges that may have affected the observed results. In the Tabletop AR condition, physical movement was hindered by obstructing chairs and tables, leading participants to avoid standing up and instead lean or hover above their seats. Additionally, a bug in the Tabletop condition, especially with T5 Mars stimuli, caused issues with the AR 'stamping' mechanism, resulting in visual movement and comments like *"Oh no, I lost Mars!"* [P23]. Future development work should address these Tabletop AR issues. In the Tangible AR condition, some participants found the interaction options and gestures less intuitive, particularly in tasks involving zoom and selection. Bimanual usage of the phone and cube simultaneously also posed challenges, leading some participants to place the cube in their lap or on the table, with one participant remarking *"I want an extra hand"* [P43]. These challenges were especially pronounced for participants with physical differences, and in one case, assistance was necessary to overcome accessibility obstacles.
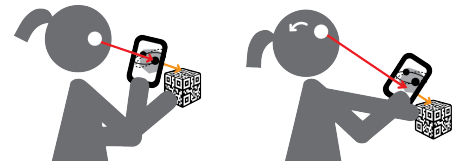

Figure 3: Diagram showing the difference in non-farsighted participants (left) versus people suspected of experiencing farsightedness (right), which requires greater viewing distance and thus restricts the range of movement for the AR tangible device.

Post-hoc analysis In a post-hoc analysis, we visually inspected the

interactions between age group in the aggregate NASA-TLX score as a measure of workload. This inspection revealed no noteworthy associations between age and workload rating. Participant numbers per categorical age group for each technology were relatively small, so we did not include a quantitative assessment of this relationship in our analysis. Qualitatively, we observed that older participants were more reluctant to stand up and walk around during the AR Tabletop condition. They also sometimes exhibited an unusual behavior in the AR Tangible condition: participants would extend their arm holding the cube all the way out, then arch their back and tip their head so as to maximize the distances between their face, their phone, and the tangible (Fig. 3). We discuss our post-hoc hypothesis for this behavior in the next section.

Thematic coding There were a number of qualitative observations about the technologies that were not captured in our quantitative results. In particular, as our participants were all in the field of astrophysics, they had some ideas for the **broader utility** of the technology in astronomy beyond figure publishing. Participants in the non-AR condition often mentioned **outreach** [P26, P40, P43, P48] and K-12 **education** [P1, P7, P41]. For example, in T3 Sundial: [P51] *"this would actually work great because I had my students make a sundial [in class] but we don't have them all day"* so a simulation would help with understanding. Several participants in the non-AR condition mentioned using the technology in the context of **planetariums** [P15, P29, P40, P50], a communication format that continues to be of great value for outreach in astronomy [12]. Participants using the AR Tactile device spoke extensively about their research topics (e.g., spectral lines [P17, P31], Active Galactic Nuclei (AGN) [P14, P31], supernovae [P20, P33] including Cassiopeia-A [P43]) and how the tangible AR would be useful for **data exploration** [P13, P47]. One person in radio astronomy analyzing 'all-sky' data noted *"my data is spherical in shape"* [P31] and saw the tangible as useful for their application due to the ability to quickly rotate with many degrees of freedom. Users of the tactile AR also mentioned education and outreach, e.g. *"I'm a TA and it's really hard when trying to explain 3D stuff like eclipses, this could be useful"* [P18]. Comments specific to astronomy were more limited in the AR Tabletop condition, although one participant exclaimed *"Ooh! I want this for my stars!"* [P26]. In both AR conditions, there were comments about how existing data visualization tools do not meet user needs. One participant [P44] mentioned frustration using Matplotlib [20], and another said they wished they could use 3D visualization more in their research but had trouble with the technology [P12].

Summary of Results The study compared non-AR technology to two AR technologies, Tangible AR and Tabletop AR, in terms of self-reported workload. Results demonstrate that the non-AR system was rated significantly lower workload than the AR systems among women. There was no significant difference among men between the AR and non-AR systems for self-reported TLX scores, but men found the Tangible AR to be lower workload on average compared to the Tabletop AR. Based on qualitative coding, we suggest that further exploration is needed to evaluate evidence of age or age-related physical differences having an effect on workload. Qualitative analysis revealed that all three technologies had benefits and challenges. Participants in the non-AR condition appeared to solve tasks more quickly, however the AR conditions lead to greater engagement as measured by extended time spent interacting with the data beyond the task duration. Participants faced challenges in the AR conditions including physical movement, difficulty in understanding the interactions, and awkwardness holding objects in both hands in AR Tangible. Although the AR technology demonstrated a 'wow' factor with increased engagement and excitement, the non-AR technology required less workload overall.

5 DISCUSSION

The results of our study (Sec. 4) exhibit **a trade-off between workload and engagement**. The quantitative results, and some of the qualitative, clearly demonstrate that the non-AR condition requires less physical and mental work. However, the qualitative observa-

tions and analysis clearly provide evidence of enjoyment and longer engagement. Future studies, in particular long-term longitudinal evaluations, will need to be conducted to better understand if increased engagement has positive benefits (e.g. information recall [7]) in spite of the increased workload. In terms of publications of astronomy journal figures, all three technologies were arguably successful and worth further investigation for the improved publishing of 3D visualizations. Additionally, while this work is focused specifically on comparisons among technologies for 3D interactive representations, this work would benefit from an extension into comparison with traditional static 2D images, such as the multiple angles shown in Bialy et. al [5].

As anticipated based on established theories in prior work, we observed **gender differences** in our quantitative analysis. Gender or biological sex differences in spatial tasks have been consistently documented in previous work where, for example, on average women do better in object location memory tasks (e.g., [38]) while men on average do better in mental rotation tasks (MRT) (e.g., [29]). Our gender-related observation could be considered as confirming the MRT-related differences (i.e., tangible AR requires mental rotation), and Ambühl & Fluri et al. have shown that MRT abilities might matter in the effective use of Merge Cube [1]. However, the non-AR version on the mobile phone should require even more mental rotation, which challenges this interpretation. Since we observed that physical differences (body composition, arm length, etc.) seem to change the user experience, a possible factor could be that the physical differences between men and women have a possible effect. Future work is merited to more carefully explore these differences for tabletop AR and AR tangibles.

Finally, the results have strong implications for **accessibility** of this technology. First, improved understanding and strategies for managing bimanual interactions is needed. We suggest exploring physical augmentations to the interaction paradigm to avoid the 'juggling' problem, such as a phone stand or something akin to electricians' soldering 'helping hands' for holding and orienting the tangible. Participants leaning away from the tangible to increase viewing distance, especially by persons wearing glasses, suggested a possible concern related to farsightedness and vision impairment. While nearsightedness is a non-issue for tangibles (participants can simply hold their phone closer to their face), farsightedness is a concern because there is a minimum viewing distance between phone and tangible, and a maximum range of motion in the arm needed for 'zooming' the cube closer or further (Fig. 3). Thus, if a user needs a greater viewing distance, they are reducing the range of movement for interaction thereby necessitating placing the cube on a surface or the phone in a stand. Finally, regardless of the technology used for any 3D journal figures, conventional accessibility guidelines should be followed, e.g. by including static images and alternative text [23].

6 CONCLUSION

We conducted a mixed-methods study to evaluate the use of non-AR, conventional AR, and AR with a tangible technologies for publishing 3D journal figures. The results of the study demonstrate that overall three technologies are usable viable options with differences between each in terms of workload (i.e., AR required more physical and mental work) and engagement (i.e., AR qualitatively was more engaging). We see significant differences by gender that require further exploration. We see great importance for the visualization community in understanding these affordances and challenges. AR is increasingly used for scientific communication, including astronomy, and we look forward to the scientific and computing communities researching together how to most effectively use this technology for publications and education, as well as for data exploration and insight.

REFERENCES

[1] A. Ambühl, L. Fluri, C. Zucker, A. Goodman, and A. Cöltekin. Effectiveness and perceived usefulness of a handheld ar cube for examining 3d spatial structures. In *SpaceCHI Workshop at the CHI2022*, 2022.

[2] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467, 2018. doi: 10.1109/TVCG.2017.2745941

[3] D. G. Barnes and C. J. Fluke. Incorporating interactive three-dimensional graphics in astronomy research papers. *New Astronomy*, 13(8):599–605, 2008.

[4] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The state of the art of spatial interfaces for 3d visualization. In *Computer Graphics Forum*, vol. 40, pp. 293–326. Wiley Online Library, 2021.

[5] S. Bialy, C. Zucker, A. Goodman, M. M. Foley, J. Alves, V. A. Semenov, R. Benjamin, R. Leike, and T. Enßlin. The per-tau shell: a giant star-forming spherical shell revealed by 3d dust observations. *The Astrophysical Journal Letters*, 919(1):L5, 2021.

[6] M. Billinghurst, H. Kato, I. Poupyrev, et al. Tangible augmented reality. *Acm siggraph asia*, 7(2):1–10, 2008.

[7] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics*, 22(1):519–528, 2015.

[8] G. Caggianese, V. Colonnese, and L. Gallo. Situated visualization in augmented reality: Exploring information seeking strategies. In *2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pp. 390–395, 2019. doi: 10.1109/SITIS.2019.00069

[9] Y.-C. Chen, H.-L. Chi, W.-H. Hung, and S.-C. Kang. Use of tangible and augmented reality models in engineering graphics courses. *Journal of Professional Issues in Engineering Education & Practice*, 137(4):267–276, 2011.

[10] A. Çöltekin, I. Lochhead, M. Madden, S. Christophe, A. Devaux, C. Pettit, O. Lock, S. Shukla, L. Herman, Z. Stachoň, et al. Extended reality in spatial sciences: A review of research challenges and future directions. *ISPRS International Journal of Geo-Information*, 9(7):439, 2020.

[11] M. Cordeil, B. Bach, A. Cunningham, B. Montoya, R. T. Smith, B. H. Thomas, and T. Dwyer. Embodied axes: Tangible, actuated interaction for 3d augmented reality data spaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376613

[12] J. K. Faherty, M. SubbaRao, R. Wyatt, A. Ynnerman, N. d. Tyson, A. Geller, M. Weber, P. Rosenfield, W. Steffen, G. Stoeckle, et al. Ideas: immersive dome experiences for accelerating science. *arXiv preprint arXiv:1907.05383*, 2019.

[13] V. Ferrer, A. Perdomo, H. R. Ali, C. Fies, and J. Quarles. Virtual humans for temperature visualization in a tangible augmented reality educational game. In *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual Augmented Reality (KELVAR)*, pp. 1–6, 2017. doi: 10.1109/KELVAR.2017.7961559

[14] P. Fleck, A. Sousa Calepso, S. Hubenschmid, M. Sedlmair, and D. Schmalstieg. Ragrug: A toolkit for situated analytics. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022. doi: 10.1109/TVCG.2022.3157058

[15] D. GmbH. Cospaces.

[16] A. A. Goodman, E. W. Rosolowsky, M. A. Borkin, J. B. Foster, M. Halle, J. Kauffmann, and J. E. Pineda. A role for self-gravity at multiple length scales in the process of star formation. *Nature*, 457(7225):63–66, 2009.

[17] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[18] A. Hassan and C. J. Fluke. Scientific visualization in astronomy: Towards the petascale astronomy era. *Publications of the Astronomical Society of Australia*, 28(2):150–170, 2011.

[19] N. R. Hedley, M. Billinghurst, L. Postner, R. May, and H. Kato. Explorations in the use of augmented reality for geographic visualization. *Presence*, 11(2):119–133, 2002.

[20] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55

[21] G. S. Inc. Glue: Multi-dimensional linked-data exploration, 2013.

[22] R. Kaehler, O. Hahn, and T. Abel. A novel approach to visualizing dark matter simulations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2078–2087, 2012.

[23] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. Accessible visualization: Design space, opportunities, and challenges. In *Computer Graphics Forum*, vol. 40, pp. 173–188. Wiley Online Library, 2021.

[24] F. Lan, M. Young, L. Anderson, A. Ynnerman, A. Bock, M. A. Borkin, A. G. Forbes, J. A. Kollmeier, and B. Wang. Visualization in astrophysics: Developing new methods, discovering our universe, and educating the earth, 2021. doi: 10.48550/ARXIV.2106.00152

[25] E. Laplace. Tulips: A tool for understanding the lives, interiors, and physics of stars. *Astronomy and Computing*, 38:100516, 2022.

[26] J. E. Lauer, E. Yhang, and S. F. Lourenco. The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin*, 145(6):537–565, June 2019. doi: 10.1037/bul0000191

[27] M. Magnor, G. Kindlmann, C. Hansen, and N. Duric. Reconstruction and visualization of planetary nebulae. *IEEE Transactions on Visualization and Computer Graphics*, 11(5):485–496, 2005.

[28] I. Merge Labs. Merge explorer (1.6.17), 2023.

[29] A. Nazareth, A. Herrera, and S. M. Pruden. Explaining sex differences in mental rotation: role of spatial activity experience. *Cognitive processing*, 14:201–204, 2013.

[30] S. M. Nizam, R. Z. Abidin, N. C. Hashim, M. C. Lam, H. Arshad, and N. Majid. A review of multimodal interaction technique in augmented reality environment. *Int. J. Adv. Sci. Eng. Inf. Technol*, 8(4-2):1460, 2018.

[31] E. Ntuli. Augmented reality in early learning: Experiences of k-3 teachers with merge cubes. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 557–560. Association for the Advancement of Computing in Education (AACE), 2019.

[32] T. O'Neill, A. Goodman, J. Soler, J. J. Han, and C. Zucker. Mapping the local bubble's magnetic field in 3d. *Authorea Preprints*, 2023.

[33] A. Strauss and J. M. Corbin. *Grounded theory in practice*. Sage, 1997.

[34] C. Swiggum, E. D'Onghia, J. Alves, J. Großschedl, M. Foley, C. Zucker, S. Meingast, B. Chen, and A. Goodman. Evidence for radial expansion at the core of the orion complex with gaia edr3. *The Astrophysical Journal*, 917(1):21, 2021.

[35] M. Taufiq, M. Nuswowati, and A. Widiyatmoko. Feasibility study of a solar system learning media based on merge cube augmented reality to embedding problem solving skills. In *Journal of Physics: Conference Series*, vol. 1918, p. 052064. IOP Publishing, 2021.

[36] M. Taufiq, M. Nuswowati, and A. Widiyatmoko. Study of the applicability level of merge cube augmented reality media on junior hight school science teachers. *Unnes Science Education Journal*, 10(3), 2021.

[37] R. N. Uppot, B. Laguna, C. J. McCarthy, G. De Novi, A. Phelps, E. Siegel, and J. Courtier. Implementing virtual and augmented reality tools for radiology education and training, communication, and clinical care. *Radiology*, 291(3):570–580, 2019.

[38] D. Voyer, A. Postma, B. Brake, and J. Imperato-McGinley. Gender differences in object location memory: A meta-analysis. *Psychonomic bulletin & review*, 14:23–38, 2007.

[39] P. Wolle, M. P. Muller, and D. Rauh. Augmented reality in scientific publications–taking the visualization of 3d structures to the next level. *ACS Chemical Biology*, 15(1):296–297, 2019.

[40] S. Zollmann, T. Langlotz, R. Grasset, W. H. Lo, S. Mori, and H. Regenbrecht. Visualization techniques in augmented reality: A taxonomy, methods and patterns. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3808–3825, 2021. doi: 10.1109/TVCG.2020.2986247

[41] C. Zucker, A. A. Goodman, J. Alves, S. Bialy, M. Foley, J. S. Speagle, J. Großschedl, D. P. Finkbeiner, A. Burkert, D. Khimey, et al. Star formation near the sun is driven by expansion of the local bubble. *Nature*, 601(7893):334–337, 2022.