APPENDIX

Appendix A: AR Concept and Differences to Tangible Brush

To describe and analyze both the original technique and our own AR-based approaches, we use Bruckner et al.'s [BIRW19] model of spatial interaction directness (Fig. 2). It classifies interactions from Class 1 (most "direct") to Class 6 (least "direct").

The lasso step of all mappings and techniques is identical. A user draws the lasso for later extrusion on the orthogonal projection on the tablet. Here, i is an inverse of v as a 2D position on the tablet is projected onto the near-clipping plane of the camera, and vice-versa. The orthogonal mapping naturally allows users to measure distances, in contrast to perspective projections. Moreover, $\mathcal{O}$ and $\mathcal{M}$ collapse due to 2D input being captured virtually on the tablet's display. We thus classify this interaction as belonging to Class 1 (i. e., direct).

Original Extrusion Interaction

In the original implementation, users move the tablet and its lasso in physical space, yet can only observe the selection effects on a 2D screen, which shows the (partially selected) dataset and the tablet's motions in two perspectively projected views. The tablet's local coordinate system (which is needed to map from $\mathcal{M}$ to $\mathcal{I}$) was reset each time the user extruded a lasso, its forward axis being defined by the physical tablet's normal during interaction and the x- and y-axes following from the screen's physical orientation. We call this mapping ***Relative-Full*** (RF). For both screens we thus have a 3D interactive space ($\mathcal{I}$ and $\mathcal{M}$ are in 3D), yet two distinct 2D output spaces $\mathcal{O}$. We analyze the spatial directness for both outputs next.

Focus on tablet: The tablet shows an orthographically projected visualization ($\mathcal{V}$). When focusing on the tablet, we consider vi to be quasi-inverse, as motions in the interaction ($\mathcal{I}$) space along the tablet's normal lead, given the near- and far-clipping planes, to the same visualization $\mathcal{V}$. om is also not the identity as the 2D $\mathcal{O}$ has to be related mentally to the 3D $\mathcal{M}$. The physical tablet, however, shows ($\mathcal{O}$) a positionally correct projection of the 3D $\mathcal{V}$ as positioned in the similarly 3D $\mathcal{M}$, helping users in their Class 5 interaction.

Focus on external screen: When users focus on the 2D screen, we consider $\mathcal{V}$ and $\mathcal{I}$ to be identical: movements of the physical and, thus, also the virtual tablet are directly mapped to translations in the visualization space. As before, om is not the identity as $\mathcal{O}$ is the 2D screen and $\mathcal{M}$ matches the physical 3D space. The lack of axis alignment between $\mathcal{O}$ and $\mathcal{M}$ in most extrusion processes can also result in a high mental workload. Based on past work [WA04, LODI16] we hypothesize that an existing or missing alignment impacts a user's mental model and the ease of creating space $\mathcal{U}$: people manipulate input devices ($\mathcal{M}$) based on what they see ($\mathcal{O}$) and their mental reference frame ($\mathcal{U}$), in particular for interactions that users need to learn. In trained interaction such as mouse input, in contrast, even though $\mathcal{M}$ (horizontal mouse) and $\mathcal{O}$ (vertical screen) do not align, users have an established mental model. We classify Tangible Brush with an external screen as Class 4.

Extrusion with the AR Tangible Brush

Next, we discuss the three AR mappings we described in Sect. 3 for the actual selection input using Brucker et al.'s model.
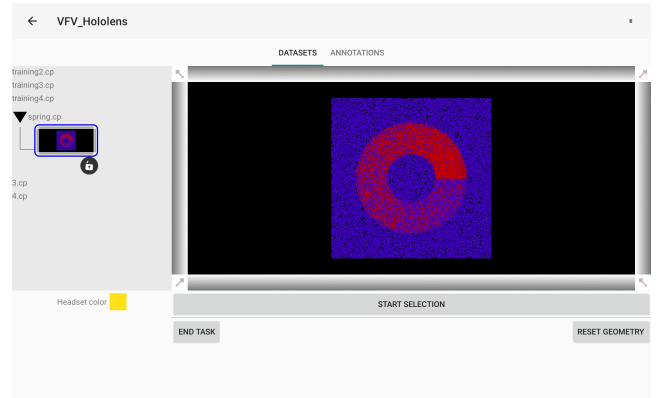
Figure 21: *The main interface of the tablet as in our overall collaborative AR application which this paper reuses. "Start Selection" shows the Tangible Brush interface (Fig. 5). "End Task" ends the current task. "Reset Geometry" replaces the dataset at its default position and orientation. The FI3D widgets surround the orthographic visualization of the dataset. In this screenshot, we show the Spring dataset as seen from its front face.*

Naïve Approach (NA)

Tablet as the focus: If the tablet is the focus, the situation is almost similar as for the traditional Tangible Brush and we fall back to classifying it as Class 5 because om is not the identity. Starting an extrusion in NA will create a "jump" for the orientation of the virtual tablet as the view will unfreeze and use the physical orientations.

AR-HMD as the focus: If the AR-HMD is the focus, $\mathcal{V}$, $\mathcal{I}$, $\mathcal{O}$, and $\mathcal{M}$ collapse, corresponding to Bruckner et al.'s Class 1.

Relative-Full (RF) and Relative-Aligned (RA)

Focus on tablet: We classify RF and RA as Class 5 because the dimensionality of $\mathcal{O}$ and $\mathcal{M}$ differ from each other. Like NA, starting an extrusion in RA creates a "jump" in the orientation of the virtual tablet, making this mapping even more spatially indirect than RF.

Focus on 3D AR view: $\mathcal{O}$ and $\mathcal{M}$ collapse for RA but do not for RF, the latter due to the difference in axis orientation. Indeed, the user redefines a new reference coordinate system for the manipulations in $\mathcal{M}$ each time they start the extrusion process using the RF mapping, compared to their (world) coordinate reference in physical space. RA, in contrast, ensures that the reference of the 3D motions in $\mathcal{M}$ matches the coordinate system of $\mathcal{O}$. In both cases vi is inverse. We thus classify RA as Class 1 and RF as Class 4.

Appendix B: Further Implementation Details

We give in this appendix more details about the implementation of our AR adaptations of Tangible Brush that we described in Sect. 4.1.

Our implementation relies on a larger ecosystem that facilitates multiple interactions necessary to explore 3D datasets. The HoloLens (FoV: $52° \times 34°$, 2500 light points per radian; weight: 566 grams) relies on Unity 3D, C#, homemade C++ libraries, and MRTK. The multi-touch tablets (display: 2560×1600 pixels, 10.5"

diagonal, 287 ppi; weight: 482 grams) relies on Android 9, Java, homemade C++ libraries, and OpenGL ES 3.0. The server relies on Linux and C/C++. We tracked the multi-touch tablet with respect to the HoloLens with the external tracking VICON system [MDB*17]. Its setup relies on fourteen VERO v2.2 (2048x1088px, submillimeter precision) placed around a cube of $6.0 \times 4.7 \times 2.0 m^3$ (surface× height). We used the VRPN communication protocol to retrieve the tracking metrics of both the HoloLens and the multi-touch tablet in real time within the VICON coordinate system.

We launch the server with the participant's ID as a parameter. The server loads the needed dataset depending on how far the participant is in the study. We use a homemade protocol on top of TCP/IP for the communication between all devices. We do not perceive any latency as we rely on a local network and do not buffer the data before sending it as Ethernet packets. We keep all relevant information on the server, to which the tablet and the HMD are paired when they connect, and they also are paired to each other via a dedicated IP entry step. The server maintains all relevant information and logs all events in JSON format, to allow us to extract metrics such as speed and accuracy using Python and R. We can thus recover from crashes of the tablet or the HMD, should these occur, without affecting the results of the study.

Datasets

Users can move and rotate 3D visualizations on the tablet via FI3D [YSI*10] (Fig. 21). In the AR conditions, the camera associated with the FI3D widgets faces the current visualization to always have it in sight. As the main view is the AR space and because the user can physically move around, parameterizing the camera based on the user' position would be unusable. In our replication of the original Tangible Brush, however, the camera associated with the FI3D widgets follows the virtual tablet position and orientation as this condition always relates to this virtual tablet that is frozen when users do not extrude anything.

We start the application with the visualizations at a fixed position. Participants can relocate the visualization vertically (y-axis) for comfortable interaction, but we locked motions along the x- and z-axes to maintain similar conditions across participants in AR. In the condition that replicates the original implementation, however, we allowed participants to move the dataset along all axes because, in the AR conditions, users can move horizontally on the ground, which acts as x- and z-translations that were impossible in the original. Rotations can only happen around the x- and y-axes, with the x-axis being defined by the user's current position in the physical space in the AR conditions. While one can argue that a z-axis rotation can be useful, we note that the drawing of the lasso can be adjusted on the tablet, replacing a z-axis rotation. We made this choice to keep all the possible interactions (rotations, moving, scaling) simple in our overall system. Finally, we allow users to reset the default positions and orientations for both the 3D dataset and the virtual tablet.

Selection Interaction Processing and Volume Selection

We send the position and orientation of the HMD in its own coordinate system to the server at each frame, at a maximum frequency of 60 Hz. The server also receives the positions and the orientations

Table 1: *Classification summarization of the different techniques and mappings using Bruckner et al.'s [BIRW19] model (Fig. 2).*

Technique	Focus	*vi*	*om*	Class	Axis aligned
Original	Tablet	quasi	no	5	**Yes**
Original	External Screen	yes	no	4	No
NA	Tablet	quasi	no	5	**Yes**
NA	HMD	yes	yes	1	**Yes**
RF	Tablet	quasi	no	5	**Yes**
RF	HMD	yes	no	4	No
RA	Tablet	quasi	no	5	**Yes**
RA	HMD	yes	yes	1	**Yes**

of both the tablet and the HMD in the coordinate system of the VICON at 60 Hz. The server then converts the tablet's coordinates into those of the HMD, the latter having been created by the HMD's private API. We then send these positions and orientations to the tablet which applies the current mapping. We then send the updated tablet's position and orientation back to the server, which passes them on to the HMD. While this design relies on several messages (VICON–Server + HoloLens–Server, Server–Tablet, Tablet–Server, Server–HoloLens), we did not perceive any latencies that would be caused by network traffic on our dedicated local network. With this design, all the decisions concerning the mappings and their implications happen on the tablet. This design allows us to change the interaction by only modifying the tablet's source code.

The chosen mapping affects the virtual tablet's orientation. Only for RF it actually differs from the physical tablet's orientation, which we compute as $o_{beg} \cdot o_{pos}^{-1} \cdot o_{cur}$ using quaternions, where o_{beg} is controlled for tangible input, o_{pos} is controlled for position input, and o_{cur} is the current tablet orientation in the user's 3D space.

Volume Selection

After placing the tablet into the 3D space, the user can draw the lasso (to be extruded) on its static orthographic view. A rapid tap removes the lasso. To optimize the computation of the volumetric shape (computational complexity of $O(n)$, n being the number of points of the lasso), we enforce a minimal distance between points on the lasso of 0.05 unit. We automatically close the lasso if its first and the last points are closer than 0.20 unit. These distance units arise from OpenGL's screen coordinate system ($-1.0 \dots +1.0$ along the screen's x- and y-axes). We compute the 3D positions of all lasso points p_i for each timestep t, filtered to a ≈ 0.05 unit resolution to improve interactivity. For timesteps $t > 1$, we display a 3D wireframe corresponding to the selection on the HMD which we update by connecting $p_{i,t}$ to $p_{i,t-1}$. We enforce that two consecutive steps, on the wireframe, are at least separated by 5 mm in the AR space to reduce clutter. We close the 3D selection volume when the user presses the "Done" button, selects another Boolean operator, ends the selection, or uses position or tangible rotation modes. We ensure this way that we do not connect endpoints that the user did not intend to connect. Updating the 3D volumes does not consume much computing power, allowing us to run it in real-time, which is a strong requirement for AR applications [Azu97, ABB*01].

Compared to the original implementation which updates the selection state of the data in real-time, users can check the intended selection volumes before they validate or invalidate them. Once vali-

dated, we compute the ROI selection on the server, which broadcasts the Boolean mask to each device. For a dataset of 160K points (i. e., the galaxy collision dataset in our study; Fig. 3), the resulting mask represents $(160 \cdot 10^3 + 7)/8 = 20$ kilobytes of data, which does not prohibit real-time interactions. When a user validates a selection, we consider all volumes and their associated Boolean operators. For each data point, we determine if it is inside or outside a given volume using raycasting by counting the intersections (even intersections for inside, odd intersections for outside) of a ray from the point to a random direction with triangles of the enclosing selection shape.

We also optimize this process with a triangle lookup in a $16 \times 16 \times 16$ grid. We then cast the ray from the point (which is at a position (i, j, k) in this 3D grid, and (x, y, z) in the 3D world), parallel to the x-axis. As a further optimization, depending on the number of triangles to consider, we either cast the ray along the $-x$ or the $+x$ direction. We apply then a ray–triangle intersection algorithm to determine the triangles the ray intersects with, filtering the triangle lists using the segmented 3D grid. We discard triangles we already considered. We further parallelized the algorithm using OpenMP to process multiple data points simultaneously.

Replicating the Original Tangible Brush Implementation

As optical see-through displays cannot render black colors, we use a dark green background color to simulate the opaque virtual screen we used to replicate the original technique. We set the virtual tablet's position by default at the front-top-left corner of the 3D cube encapsulating the data, facing the center of the cube. With this choice, we do not align the mathematically generated datasets that have axis-aligned features with the default tablet's position and orientation (along x-, y-, and z-axis). As in the original, on the virtual screen's left side we render a perspective view from the point of view of the current virtual tablet. Its field-of-view (fov) depends on the current size of the virtual tablet: $verticalFOV = atan(3 \cdot sizeX)$, with $verticalFOV$ in radian and $sizeX$ (the virtual tablet's width) in meter. On the virtual screen's right side, again as in the original, we render the scene using a birds-eye-view camera which we placed at $(0, 1\,m, -1.75\,m)$ and which faces the center of the scene, where we place the dataset by default.

As in the original technique, we apply translations with respect to the tablet's coordinate system, with the z-axis defined by the normal of the virtual tablet. This control is convenient when looking at either the tablet or the virtual screen's left side, as the axes of these respective cameras are aligned with the translation motion. Based on a pilot study, we allow participants to rotate the 3D scene around the center of the dataset using the tablet as a tangible device. 3D rotations are strong depth cues that allow users to align both the tablet and the dataset along an axis they can understand, and to review the 3D extruded-wireframed volumes within the 3D scene.

Appendix C: Experiments

In Sect. 4 we presented our experiment designs and described how we ran the two studies. In the interest of full disclosure, here we detail the issues we ran into while running the experiments. Those issues are minor and did not affect our results. In contrast, we explained all major issues in our limitations section (Sect. 7).
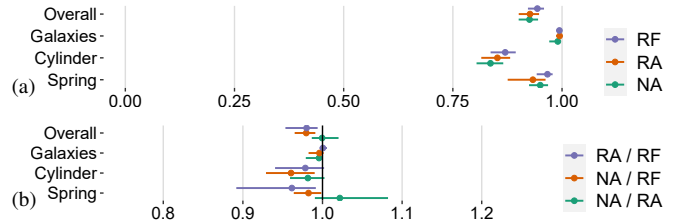

Figure 22: *Raw data (a) and pairwise within-subjects comparisons (b) of the F1 accuracy score, 1st experiment.*
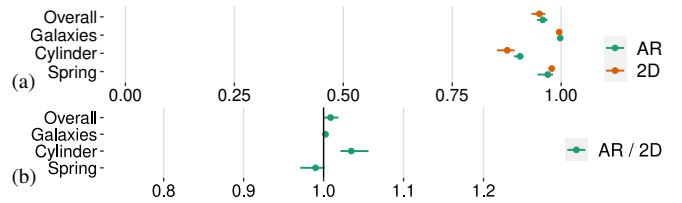

Figure 23: *Raw data (a) and pairwise within-subjects comparisons (b) of the F1 accuracy score, 2nd experiment.*

Due to issues with the HoloLens (e. g., opening the startup menu or losing itself in the physical space), we adjusted two trials of the first study to remove the time we lost to fix the system, removing 90 and 30 seconds for $(P_{ID}; Technique; Dataset; SubTrial) = (5; NA; Spring; 0)$ and $(7; NA; Cylinder; 0)$, respectively, whose durations we recovered from our video recordings. As the HoloLens sometimes went into its saving power mode during breaks, potential errors could appear, e. g., creating ghost TCP/IP connections. Some participants also ask for additional breaks while we did not create a functionality to pause the experiment. We thus rebooted the environment and start again from where the participant was when he or she was ready, resulting in two sub-log files instead of one. These reboots concern P_1, P_3, and P_{11} for the first experiment and for participants P_{1-2}, P_6, P_8, P_{14}, and P_{17} for the second experiment. For P_{14} of the second experiment, especially, the HoloLens' startup menu showed up too frequently when he manipulated the tablet. Another gesture to open that menu may be needed.

To be fully transparent we share in this appendix the F1 metrics (Fig. 22 and Fig. 23) that we did not present in the main part of the paper. The conclusions we drew from these figures are exactly the same as those that we obtained by analyzing the MCC metrics (the accuracy metric) for the two reported experiments.

We also share in Table 2 and in Table 8 all the logged numerical values of our raw data statistical analysis (i. e., considering the whole population), in Table 3 and in Table 9 the post-analysis of determining the learning effects on participants following their technique order, and in Table 4 and in Table 10 all the numerical values of our raw data statistical analysis of the NASA-TLX questionnaire. Finally, all pairwise comparisons we performed can be found in Table 5 and Table 11 for the quantitative metrics, in Table 6 and Table 12 for the learning effects on the time-completion task, and in Table 7 and Table 13 for the NASA-TLX qualitative questionnaire. Those tables about within-subject pairwise comparisons give a 95% confidence intervals of the effect sizes between each pair of conditions we studied.

Table 2: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the logged quantitative raw metrics we gathered in the first experiment. Time-Completion Tasks (TCT) are in seconds. "Constraint" depicts the proportion of constraint operations that contributed to the final results over unconstrained ones.*

Metric	Technique	Mean	Min CI	Max CI	SD
Overall					
MCC	NA	0.88	0.85	0.91	0.11
MCC	RA	0.89	0.85	0.92	0.12
MCC	RF	0.91	0.89	0.94	0.09
F1	NA	0.93	0.90	0.94	0.08
F1	RA	0.93	0.90	0.95	0.09
F1	RF	0.94	0.92	0.96	0.07
TCT	NA	180	170	210	79
TCT	RA	170	140	200	130
TCT	RF	190	160	220	140
Constraint	NA	0.79	0.67	0.87	0.39
Constraint	RA	0.86	0.75	0.93	0.32
Constraint	RF	0.93	0.84	0.97	0.22
Galaxies					
MCC	NA	0.98	0.95	0.99	0.03
MCC	RA	0.99	0.98	0.99	0.01
MCC	RF	0.99	0.98	0.99	0.02
F1	NA	0.99	0.97	1.00	0.02
F1	RA	0.99	0.99	1.00	0.01
F1	RF	0.99	0.99	1.00	0.01
TCT	NA	170	140	200	67
TCT	RA	170	140	200	69
TCT	RF	170	130	220	110
Constraint	NA	0.81	0.59	0.93	0.35
Constraint	RA	0.87	0.65	0.97	0.30
Constraint	RF	0.94	0.83	0.98	0.14
Cylinder					
MCC	NA	0.79	0.75	0.83	0.09
MCC	RA	0.81	0.76	0.85	0.09
MCC	RF	0.83	0.79	0.86	0.08
F1	NA	0.84	0.80	0.86	0.07
F1	RA	0.85	0.81	0.88	0.07
F1	RF	0.87	0.84	0.89	0.06
TCT	NA	150	130	170	44
TCT	RA	110	85	150	86
TCT	RF	150	115	180	83
Constraint	NA	0.83	0.51	0.94	0.38
Constraint	RA	0.94	0.67	1.00	0.24
Constraint	RF	1.00	1.00	1.00	0.00
Spring					
MCC	NA	0.88	0.83	0.92	0.098
MCC	RA	0.87	0.78	0.91	0.14
MCC	RF	0.92	0.88	0.95	0.071
F1	NA	0.95	0.92	0.97	0.047
F1	RA	0.93	0.88	0.96	0.087
F1	RF	0.97	0.94	0.98	0.036
TCT	NA	250	220	290	79
TCT	RA	260	200	330	150
TCT	RF	290	230	360	150
Constraint	NA	0.72	0.46	0.86	0.43
Constraint	RA	0.75	0.54	0.89	0.39
Constraint	RF	0.84	0.62	0.96	0.34

Table 3: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of the time-completion task (in seconds) depending on the participants' personalized technique order; first experiment.*

Technique	Mean	Min CI	Max CI	SD
Overall				
First	220	200	260	62
Second	170	150	200	51
Third	160	140	180	40
Galaxies				
First	220	190	270	86
Second	140	120	180	68
Third	140	120	170	55
Cylinder				
First	150	120	200	97
Second	140	110	180	77
Third	120	96	140	53
Spring				
First	320	260	400	160
Second	250	200	300	120
Third	240	200	290	91

Table 4: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the Nasa-TLX subscales we gathered in the first experiment.*

Metric	Technique	Mean	Min CI	Max CI	SD
Effort	NA	10.5	8.6	12.5	4.2
Effort	RA	9.2	7.3	10.9	4.1
Effort	RF	10.6	8.6	11.9	3.5
Frustration	NA	7.0	4.7	9.3	5.2
Frustration	RA	5.2	3.5	7.3	4.1
Frustration	RF	7.1	4.8	9.1	4.7
Temporal	NA	6.9	4.9	8.6	4.1
Temporal	RA	7.4	5.4	9.3	4.2
Temporal	RF	7.7	5.8	9.3	3.9
Physical	NA	10.8	8.2	13.1	5.3
Physical	RA	9.4	7.2	11.7	5.0
Physical	RF	8.8	6.7	11.2	5.1
Mental	NA	9.7	7.8	12.1	4.7
Mental	RA	9.7	7.3	11.8	5.0
Mental	RF	10.0	8.1	11.9	4.3
Performance	NA	13.1	11.4	14.6	3.5
Performance	RA	16.1	14.8	17.1	2.5
Performance	RF	15.1	13.6	16.2	2.8

Table 5: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the logged quantitative metrics we gathered in the first experiment. Statistics of the raw data are in Table 2.*

Metric	Conditions	Mean	Min CI	Max CI	SD
Overall					
MCC	NA-RA (Dif.)	>-0.01	-0.02	0.02	0.08
MCC	NA-RF (Dif.)	-0.03	-0.05	-0.01	0.06
MCC	RA-RF (Dif.)	-0.03	-0.05	-0.01	0.08
F1	NA/RA (Ratio)	1.00	0.99	1.02	0.07
F1	NA/RF (Ratio)	0.98	0.97	0.99	0.05
F1	RA/RF (Ratio)	0.98	0.95	0.99	0.06
TCT	NA/RA (Ratio)	1.10	0.93	1.29	0.83
TCT	NA/RF (Ratio)	0.96	0.83	1.11	0.62
TCT	RA/RF (Ratio)	0.88	0.74	1.03	0.69
Constraint	NA-RA (Dif.)	-0.07	-0.17	0.02	0.37
Constraint	NA-RF (Dif.)	-0.14	-0.26	-0.04	0.42
Constraint	RA-RF (Dif.)	-0.07	-0.17	-0.01	0.29
Galaxies					
MCC	NA-RA (Dif.)	-0.01	-0.03	<0.01	0.03
MCC	NA-RF (Dif.)	-0.01	-0.03	<0.01	0.03
MCC	RA-RF (Dif.)	<0.01	>-0.01	0.01	0.01
F1	NA/RA (Ratio)	1.00	0.98	1.00	0.02
F1	NA/RF (Ratio)	1.00	0.98	1.00	0.02
F1	RA/RF (Ratio)	1.00	1.0	1.01	0.01
TCT	NA/RA (Ratio)	1.00	0.79	1.27	0.57
TCT	NA/RF (Ratio)	0.98	0.77	1.26	0.59
TCT	RA/RF (Ratio)	0.98	0.75	1.29	0.69
Constraint	NA-RA (Dif.)	-0.06	-0.26	0.05	0.30
Constraint	NA-RF (Dif.)	-0.12	-0.33	0.01	0.36
Constraint	RA-RF (Dif.)	-0.07	-0.27	0.01	0.26
Cylinder					
MCC	NA-RA (Dif.)	-0.02	-0.05	<0.01	0.05
MCC	NA-RF (Dif.)	-0.04	-0.08	-0.01	0.08
MCC	RA-RF (Dif.)	-0.02	-0.06	<0.01	0.07
F1	NA/RA (Ratio)	0.98	0.96	1.00	0.05
F1	NA/RF (Ratio)	0.96	0.93	0.99	0.07
F1	RA/RF (Ratio)	0.98	0.94	1.00	0.06
TCT	NA/RA (Ratio)	1.33	0.94	1.87	1.30
TCT	NA/RF (Ratio)	1.04	0.79	1.37	0.74
TCT	RA/RF (Ratio)	0.78	0.56	1.09	0.74
Constraint	NA-RA (Dif.)	-0.11	-0.50	-0.06	0.32
Constraint	NA-RF (Dif.)	-0.17	-0.50	-0.06	0.38
Constraint	RA-RF (Dif.)	-0.06	-0.28	0.00	0.24
Spring					
MCC	NA-RA (Dif.)	0.02	-0.03	0.08	0.12
MCC	NA-RF (Dif.)	-0.04	-0.07	-0.01	0.07
MCC	RA-RF (Dif.)	-0.06	-0.14	-0.02	0.12
F1	NA/RA (Ratio)	1.02	0.99	1.08	0.10
F1	NA/RF (Ratio)	0.98	0.96	1.00	0.04
F1	RA/RF (Ratio)	0.96	0.89	0.99	0.09
TCT	NA/RA (Ratio)	0.99	0.75	1.31	0.69
TCT	NA/RF (Ratio)	0.87	0.67	1.13	0.57
TCT	RA/RF (Ratio)	0.87	0.66	1.16	0.64
Constraint	NA-RA (Dif.)	-0.04	-0.26	0.16	0.47
Constraint	NA-RF (Dif.)	-0.13	-0.38	0.09	0.52
Constraint	RA-RF (Dif.)	-0.09	-0.31	0.04	0.36

Table 6: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the time-completion task (in seconds) depending on the participants' personalized technique order; first experiment. Statistics of the raw data are in Table 3*

Condition Order (Ratio)	Mean	Min CI	Max CI	SD
Overall				
First/Second	1.31	1.12	1.53	0.44
First/Third	1.41	1.16	1.72	0.63
Second/Third	1.08	0.94	1.23	0.31
Galaxies				
First/Second	1.55	1.32	1.82	0.54
First/Third	1.58	1.33	1.88	0.60
Second/Third	1.02	0.85	1.22	0.41
Cylinder				
First/Second	1.11	0.79	1.56	1.07
First/Third	1.33	0.94	1.88	1.34
Second/Third	1.19	0.90	1.57	0.84
Spring				
First/Second	1.29	1.02	1.63	0.72
First/Third	1.32	0.97	1.81	1.12
Second/Third	1.03	0.83	1.28	0.52

Table 7: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the Nasa-TLX subscales we gathered in the first experiment. Statistics of the raw data are in Table 4*

Metric	Conditions (Ratio)	Mean	Min CI	Max CI	SD
Effort	NA/RA	1.16	0.91	1.41	0.56
Effort	NA/RF	0.99	0.83	1.25	0.62
Effort	RA/RF	0.85	0.67	1.01	0.34
Frustration	NA/RA	1.25	0.96	1.57	0.67
Frustration	NA/RF	0.91	0.66	1.30	0.96
Frustration	RA/RF	0.73	0.51	1.05	0.81
Temporal	NA/RA	0.91	0.72	1.14	0.58
Temporal	NA/RF	0.85	0.56	1.15	0.66
Temporal	RA/RF	0.94	0.66	1.15	0.51
Physical	NA/RA	1.09	0.87	1.32	0.52
Physical	NA/RF	1.18	0.93	1.55	0.97
Physical	RA/RF	1.08	0.97	1.22	0.30
Mental	NA/RA	1.06	0.81	1.40	0.77
Mental	NA/RF	0.98	0.73	1.27	0.71
Mental	RA/RF	0.91	0.70	1.22	0.80
Performance	NA/RA	0.81	0.70	0.91	0.22
Performance	NA/RF	0.87	0.75	0.95	0.20
Performance	RA/RF	1.07	1.00	1.17	0.21

Table 8: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the logged quantitative raw metrics we gathered in the second experiment. Time-Completion Tasks (TCT) are in seconds. "Constraint" depicts the proportion of constraint operations that contributed to the final results over unconstrained ones.*

Metric	Technique	Mean	Min CI	Max CI	SD
Overall					
MCC	AR	0.93	0.91	0.95	0.07
MCC	2D	0.93	0.90	0.94	0.07
F1	AR	0.96	0.94	0.97	0.05
F1	2D	0.95	0.93	0.96	0.06
TCT	AR	250	220	290	160
TCT	2D	240	210	290	200
Constraint	AR	0.95	0.90	0.98	0.14
Constraint	2D	0.97	0.95	0.98	0.06
Galaxies					
MCC	AR	1.00	0.99	1.00	<0.01
MCC	2D	0.99	0.98	0.99	0.01
F1	AR	1.00	1.00	1.00	<0.01
F1	2D	1.00	0.99	1.0	0.01
TCT	AR	210	170	270	110
TCT	2D	200	140	280	190
Constraint	AR	0.97	0.92	0.99	0.07
Constraint	2D	0.94	0.89	0.97	0.09
Cylinder					
MCC	AR	0.88	0.86	0.89	0.03
MCC	2D	0.84	0.81	0.86	0.06
F1	AR	0.91	0.89	0.91	0.02
F1	2D	0.88	0.85	0.89	0.05
TCT	AR	200	160	240	96
TCT	2D	190	160	220	63
Constraint	AR	1.00	1.00	1.00	0.00
Constraint	2D	1.00	0.99	1.00	0.01
Spring					
MCC	AR	0.93	0.88	0.95	0.07
MCC	2D	0.95	0.93	0.96	0.03
F1	AR	0.97	0.95	0.98	0.04
F1	2D	0.98	0.97	0.98	0.01
TCT	AR	380	310	470	170
TCT	2D	400	310	510	250
Constraint	AR	0.88	0.75	0.95	0.21
Constraint	2D	0.97	0.95	0.99	0.04

Table 9: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of the time-completion task (in seconds) depending on the participants' personalized technique order; second experiment.*

Technique	Mean	Min CI	Max CI	SD
Overall				
First	290	240	350	130
Second	210	180	250	71
Galaxies				
First	250	190	340	190
Second	170	130	210	95
Cylinder				
First	220	180	260	89
Second	170	140	200	63
Spring				
First	440	340	570	280
Second	340	290	410	130

Table 10: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the Nasa-TLX subscales we gathered in the second experiment.*

Metric	Technique	Mean	Min CI	Max CI	SD
Effort	AR	8.6	6.4	10.8	4.8
Effort	2D	12.8	10.8	14.8	4.4
Frustration	AR	4.6	2.6	7.5	5.2
Frustration	2D	9.2	6.4	12.2	6.4
Temporal	AR	6.4	4.5	8.9	4.9
Temporal	2D	7.9	5.4	10.4	5.7
Physical	AR	9.7	7.7	11.3	4.1
Physical	2D	8.4	5.8	11.3	6.4
Mental	AR	9.8	6.9	12.2	5.8
Mental	2D	11.7	8.8	13.9	5.7
Performance	AR	16.5	14.1	17.6	3.4
Performance	2D	13.2	11.1	14.8	4.0

Table 11: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the logged quantitative metrics we gathered in the second experiment. Statistics of the raw data are in Table 8.*

Metric	Conditions	Mean	Min CI	Max CI	SD
Overall					
MCC	AR-2D (Dif.)	0.0084	-0.0046	0.020	0.046
F1	AR/2D (Ratio)	1.0	1.0	1.0	0.034
TCT	AR/2D (Ratio)	1.0	0.89	1.2	0.67
Constraint	AR-2D (Dif.)	-0.021	-0.068	0.0092	0.14
Galaxies					
MCC	AR-2D (Dif.)	0.0047	0.0016	0.011	0.0092
F1	AR/2D (Ratio)	1.0	1.0	1.0	0.0046
TCT	AR/2D (Ratio)	1.1	0.80	1.5	0.86
Constraint	AR-2D (Dif.)	0.032	-0.0040	0.071	0.083
Cylinder					
MCC	AR-2D (Dif.)	0.037	0.023	0.055	0.036
F1	AR/2D (Ratio)	1.0	1.0	1.1	0.038
TCT	AR/2D (Ratio)	1.0	0.81	1.3	0.64
Constraint	AR-2D (Dif.)	0.0023	0.0	0.0069	0.0098
Spring					
MCC	AR-2D (Dif.)	-0.016	-0.050	0.0067	0.061
F1	AR/2D (Ratio)	0.99	0.97	1.0	0.033
TCT	AR/2D (Ratio)	0.97	0.76	1.2	0.57
Constraint	AR-2D (Dif.)	-0.096	-0.22	-0.020	0.21

Table 12: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the time-completion task (in seconds) depending on the participants' personalized technique order; second experiment. Statistics of the raw data are in Table 9*

Condition Order (Ratio)	Mean	Min CI	Max CI	SD
Overall				
First/Second	1.35	1.15	1.59	0.47
Galaxies				
First/Second	1.50	1.21	1.87	0.77
Cylinder				
First/Second	1.27	1.02	1.58	0.65
Spring				
First/Second	1.29	1.05	1.59	0.61

Table 13: *Mean, Minimum 95% Confidence Intervals, Maximum 95% Confidence Intervals, and Standard Deviation of all the within-subject pairwise comparisons of the Nasa-TLX subscales we gathered in the second experiment. Statistics of the raw data are in Table 10*

Metric	Conditions (Ratio)	Mean	Min CI	Max CI	SD
Effort	AR/2D	0.63	0.50	0.80	0.36
Frustration	AR/2D	0.48	0.37	0.64	0.41
Temporal	AR/2D	0.88	0.54	1.25	0.96
Physical	AR/2D	1.39	1.04	2.24	3.12
Mental	AR/2D	0.81	0.70	0.91	0.24
Performance	AR/2D	1.25	1.14	1.45	0.39