Revisiting Milgram and Kishino's Reality-Virtuality Continuum

Richard Skarbez[1]*[†], Missie Smith[2][†] and Mary C. Whitton[3]

[1] *Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC, Australia,*
[2] *Independent Researcher, Detroit, MI, United States,* [3] *Department of Computer Science, University of North Carolina, Chapel Hill, NC, United States*

Since its introduction in 1994, Milgram and Kishino's reality-virtuality (RV) continuum has been used to frame virtual and augmented reality research and development. While originally, the RV continuum and the three dimensions of the supporting taxonomy (extent of world knowledge, reproduction fidelity, and extent of presence metaphor) were intended to characterize the capabilities of visual display technology, researchers have embraced the RV continuum while largely ignoring the taxonomy. Considering the leaps in technology made over the last 25 years, revisiting the RV continuum and taxonomy is timely. In reexamining Milgram and Kishino's ideas, we realized, first, that the RV continuum is actually discontinuous; perfect virtual reality cannot be reached. Secondly, mixed reality is broader than previously believed, and, in fact, encompasses conventional virtual reality experiences. Finally, our revised taxonomy adds coherence, accounting for the role of users, which is critical to assessing modern mixed reality experiences. The 3D space created by our taxonomy incorporates familiar constructs such as presence and immersion, and also proposes new constructs that may be important as mixed reality technology matures.

Keywords: virtual reality, augmented reality, mixed reality, presence, immersion, coherence, taxonomy

1. INTRODUCTION

In 1994, Paul Milgram and Fumio Kishino published "A Taxonomy of Mixed Reality Visual Displays," simultaneously introducing to the literature the notion of the reality-virtuality (RV) continuum and the term "mixed reality" (MR) (Milgram and Kishino, 1994). In the succeeding quarter century, this work has been cited thousands of times, cementing it as one of the seminal works in our field (A related paper by Milgram, Haruo Takemura, Akira Utsumi, and Kishino titled, "Augmented reality: A class of displays on the reality-virtuality continuum," appeared later that year, and also has thousands of citations, Milgram et al., 1994). In that same quarter century, our field has rapidly evolved. Films like *Minority Report*, *Iron Man*, and *Ready Player One* have firmly established augmented reality (AR) and virtual reality (VR) in popular culture. At the same time, AR and VR technologies have rapidly become higher quality, cheaper, and more widely available. As a result, millions of consumers now have access to AR experiences on their mobile phones (e.g., *Pokémon GO*), or VR experiences on the Facebook Oculus or HTC Vive head-mounted displays (e.g., *Beat Saber*). In light of this rapid technological evolution, we believe it is worth revisiting core concepts such as the reality-virtuality continuum.

In this article, we reflect on the RV continuum, the meaning of "mixed reality," and Milgram and Kishino's taxonomy of MR display devices. That reflection leads us to three main points. First, we argue that the RV continuum is, in fact, discontinuous: that the "virtual reality" endpoint is

unreachable, and any form of technology-mediated reality is, in fact, mixed reality. Second, we consider the term "mixed reality," and argue for the continuing utility of Milgram and Kishino's definition, with one small but significant change: Instead of requiring that real and virtual objects be combined within a single display, we propose that real and virtual objects and stimuli could be combined within a single percept. Finally, we present a taxonomy—inspired by the taxonomy in Milgram and Kishino's original paper—that can categorize users' mixed reality experiences, and discuss some of the implications of this taxonomy.

We choose to focus our discussion specifically on the concepts and constructs introduced by Milgram and Kishino in their original papers. We make this choice in the interest of clarity and readability. However, we acknowledge that we are not the first or the only authors to expand upon their work in the last 20 years. Koleva, Benford, and Greenhalgh explored the idea of *boundaries* between physical and virtual spaces in mixed reality environments, and delineated some of their properties (Koleva et al., 1999). Lindeman and Nova proposed a classification framework for multisensory AR experiences based on where the real and virtual stimuli are mixed (Lindeman and Noma, 2007). Normand, Servières, and Moreau reviewed existing taxonomies of AR applications and proposed their own (Normand et al., 2012). Barba, MacIntyre, and Mynatt argued for using a definition of MR inspired by Mackay (Mackay, 1998) and a definition of AR from Azuma (Azuma, 1997), and used the RV continuum to describe the relationship between the two (Barba et al., 2012). Mann and colleagues discussed a variety of "realities"–virtual, augmented, mixed, and mediated–and proposed *multimediated* reality (Mann et al., 2018). Speiginer and MacIntyre introduced the concept of *reality layers* and proposed the *Environment-Augmentation framework* for reasoning about mixed reality applications (Speiginer and MacIntyre, 2018). Speicher, Hall, and Nebeling investigated the definition of mixed reality through a literature review and a series of interviews with domain experts. They also proposed a conceptual framework for MR (Speicher et al., 2019). Of note is that while these papers enrich the discussion regarding mixed reality, none challenges the central notion of the RV continuum, nor do they generally propose alternative definitions of mixed reality.

2. REVISITING THE REALITY-VIRTUALITY CONTINUUM

The RV continuum, as initially proposed by Milgram and Kishino, is shown in **Figure 1**. They anchor one end with a purely real environment, "consisting solely of real objects," and the other, with a purely virtual environment, "consisting solely of virtual objects" (Milgram and Kishino, 1994). They consider any environment which consists of a blending of real and virtual objects to be mixed reality (MR). Mixed reality environments where the real world is augmented with virtual content are called augmented reality (AR), while those where most of the content is virtual but there is some awareness or inclusion of real world objects are called augmented virtuality (AV). Of note is that this

FIGURE 1 | Milgram and Kishino's reality-virtuality continuum (adapted from Milgram et al., 1994).
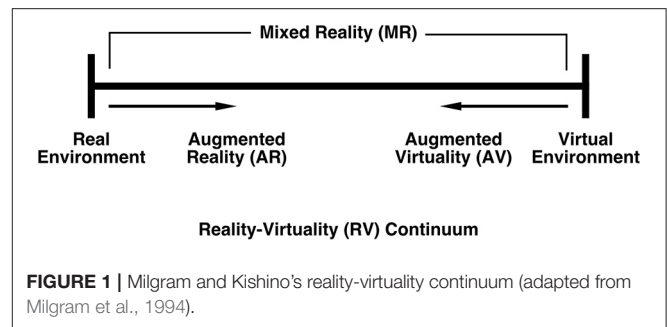
original version of the continuum was explicitly concerned only with visual displays.

While the original version of the continuum has undoubtedly served the field well, we have identified limitations. One is that, as mentioned above, Milgram and Kishino were explicitly concerned with visual displays, and primarily with display hardware. A second is that nowhere in this continuum is seen the notion of an observer or a user with senses other than visual and prior life experiences. Finally, content was described only in relation to realism (e.g., wireframes vs. 3D renderings), with no concern for the coherence of the overall experience. We will soon argue that the notion of an environment without an experiencing being—the aforementioned observer—is incomplete. That is, the mediating technology, content conveyed, and resulting user experience must be considered together to adequately describe MR experiences.

The first limitation is fairly straightforward, and in fact was commented upon by Milgram and Kishino in their original paper: "It is important to point out that, although we focus in this paper exclusively on mixed reality visual displays, many of the concepts proposed here pertain as well to analogous issues associated with other display modalities" (Milgram and Kishino, 1994). In our revisiting of the RV continuum we have taken into consideration the advances in synthesizing and displaying data for the multiple senses.

Today's processor speeds make it possible to deliver high quality *audio signals*, for instance, by modeling room acoustics (Savioja and Svensson, 2015) and sound propagation in multi-room spaces (Liu and Manocha, 2020).

Haptic displays mimic solid surfaces and other tactile stimuli. Haptics can be *active*, with solid surfaces approximated with forces supplied by a device (Salisbury and Srinivasan, 1997), or *passive* where the user feels real objects that correspond to virtual objects (Insko, 2001; Azmandian et al., 2016).

Heilig's 1962 Sensorama (Heilig, 1962) presaged integration of *scent* into virtual reality systems (Yanagida, 2012). Olfactory interfaces have matured to the point that recent work by Flavián, Ibáñez-Sánchez, and Orús explored how to make olfactory input *more effective*, rather than simply focusing on making it work (Flavián et al., 2021).

The complete *taste* experience combines sound, smell, haptics, and a chemical substance that mimics natural taste and simulates the taste buds. The Food Simulator project (Iwata et al., 2004) tackled the haptic component of taste, and recent work reports
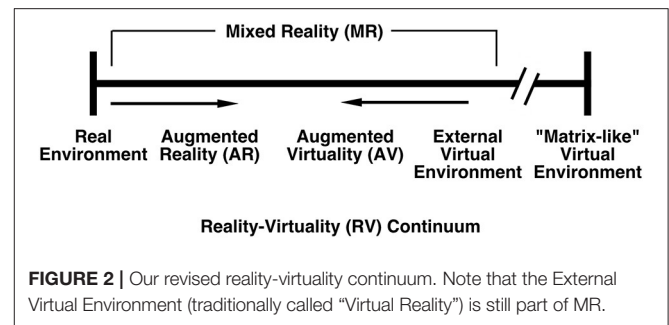
on a taste display that synthesizes and delivers tastes that match those sampled with a taste sensor (Miyashita, 2020).

All that is to say that researchers have now demonstrated at least preliminary abilities to deliver computer-generated stimuli to all the exteroceptive senses—that is, those senses responding to stimuli that come from outside the body. As progress continues, we may approach the capabilities of Ivan Sutherland's Ultimate Display—"a room within which the computer can control the existence of matter" (Sutherland, 1965) (In popular culture, one can see the Ultimate Display in the Holodecks of the *Star Trek* franchise).

However, we argue that even if we were to have the Ultimate Display, it would still fall within the realm of mixed reality. That is because, even with total control of a user's exteroceptive senses, we still would not have control over their *interoceptive senses*—the senses that monitor the body's internal state, such as the vestibular and proprioceptive senses. And even in the Ultimate Display, there would be circumstances in which these interoceptive senses would be in conflict with the information being supplied to the exteroceptive senses. For example, consider how you, as a user, might feel if the Ultimate Display were used to generate a virtual environment depicting a spacewalk. The visual display could be completely indistinguishable from the real thing, but you would still know which direction *down* was, and your feet would still be on the floor. We argue that these sensory conflicts are *inherent* to conventional virtual reality systems (which we refer to as *External Virtual Environments* in **Figure 2**). Observers experience external virtual environments through stimulation of the five basic exteroceptive senses (i.e., sight, hearing, touch, smell, and taste) while interoceptive senses remain unaltered. An important characteristic of external virtual environments is that they are *unable* to manipulate interoceptive senses.

There is, however, a popular conception of a "virtual environment" in which these sensory conflicts could be avoided: the Matrix, from the popular film series of the same name. In the *Matrix* films, sensory agreement is accomplished by direct brain stimulation: a person's sensory organs are in some way disconnected from their brain such that both interoceptive (e.g., proprioception) and exteroceptive (e.g., sight) senses are stimulated by technology. We argue that this is the only type of virtual environment that could exist outside of the mixed reality spectrum. Every other system, even the Ultimate Display, presents mixed—and potentially conflicting—exteroceptive and/or interoceptive stimuli to the user. Following this logic, we present our revised RV continuum seen in **Figure 2**, which, on the right end, includes a discontinuity between external virtual environments and the right-end anchor, "Matrix-like" VR.

We feel that the virtual environment endpoint in the original continuum was ill-defined, being any environment "consisting solely of virtual objects," although it was implied that such an environment "is one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world" (Milgram and Kishino, 1994). Most subsequent authors seem to have assumed that the virtual environment endpoint comprises what we have called external VEs, but, as we have argued, these are never "totally immers[ive]" or "completely synthetic" because they cannot

FIGURE 2 | Our revised reality-virtuality continuum. Note that the External Virtual Environment (traditionally called "Virtual Reality") is still part of MR.

control or manipulate the interoceptive senses. Furthermore, the display devices in such external VEs *are themselves real objects*, situated in the real environment. As a result, users experience such external VEs as mixed reality, with virtual objects situated within a real environment. The discontinuity in our revised continuum makes it explicit that there are real and substantial differences between external virtual environments and "Matrix-like" virtual environments.

3. THE MEANING OF MIXED REALITY

"Within this [reality-virtuality] framework it is straightforward to define a generic Mixed Reality (MR) environment as one in which real world and virtual world objects are presented together within a single display" (Milgram et al., 1994).

The preceding quote clearly defines MR, at least as Milgram and his colleagues envisioned it. MR is any display (interpreted broadly) that presents a combination of real and virtual objects that are perceived at the same time. This can be achieved in a variety of ways. Virtual objects can be visually overlaid on the real world, using optical- or video-see-through display techniques. Alternatively, real world content can be integrated into a virtual world by embedding a live video stream or, appealing to a different sense, by incorporating tracked haptic objects into a virtual experience.

Since Milgram and Kishino's initial publication, researchers have arrived at vastly different and sometimes conflicting definitions of MR. For example, MR has been defined as a combination of AR and VR, as a synonym for AR, as a "stronger" version of AR, or as Milgram and Kishino defined it (Speicher et al., 2019). In popular culture, the distinction between augmented and mixed reality has also been blurred, with some companies such as Intel[1] describing mixed reality as spatially-located and interactive with the real world, while augmented reality specifically does not include interaction. Microsoft[2] defines augmented reality as the overlaying of graphics onto video—such as AR presented on mobile phones or tablets—while mixed reality requires a combination of the physical and the

---

[1]https://www.intel.com/content/www/us/en/tech-tips-and-tricks/virtual-reality-vs-augmented-reality.html

[2]https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality

virtual. An example is the Microsoft HoloLens game RoboRaid[3], in which enemies seem to exist *on* the walls and can be occluded by real objects in the real room in which the game is being played; if you move to a different room, the enemies' locations adapt to the new physical configuration. A commonly-employed shorthand is that MR systems possess knowledge about the physical world, while AR systems do not (The notion of world knowledge is discussed at length in the following section).

We propose to unify these various definitions by making a small but fundamental change to Milgram et al.'s original definition of mixed reality. This change addresses the second limitation noted in section 2, i.e., the missing user/observer. To account for the importance of how the real or virtual content is *observed*, we propose this definition: a mixed reality (MR) environment is one in which real world and virtual world objects and stimuli are presented together within a single *percept*. That is, when a user simultaneously perceives both real and virtual content, *including across different senses*, that user is experiencing mixed reality. As such, our definition agrees with Milgram et al.'s original assertion that augmented reality is a subset of mixed reality. However, we argue that external virtual reality, what some consider to be the end point of the original RV continuum, is *also* a subset of mixed reality, because an individual may perceive virtual content with some senses and real content with others (including interoception). For example, simulating eating a meal by applying the most sophisticated visual, audio, haptic, olfactory, and taste cues may be convincing to a user, but at some point they would likely realize that they are still not satiated, and in fact, may be more hungry than when they began. This conflict between exteroception and interoception shows how conflicting signals in a single percept can make an experience incongruent. It is for this reason that there is a discontinuity on our revised continuum, because true virtual reality exists only when all senses—exteroceptive and interoceptive—are fully overridden by computer-generated content.

We acknowledge the potential criticism that our broaddefinition that includes external virtual environments makes "mixed reality" too inclusive and potentially confusing. (In the words of Speicher, Hall, and Nebeling's interviewees, "if [a console video game] is MR, then everything is" Speicher et al., 2019). Milgram and Kishino's original definition required the *(visually) displayed content* to be a mixture of real and virtual, while our proposed redefinition merely requires the user's *overall sensory experience, the percept,* be a mixture of real and virtual. Our response to the criticism that our definition of MR is too broad is two-fold. First, as illustrated earlier in this section, the many definitions of MR were already a source of confusion. Second, as we discuss in the following section, "mixed reality" is not intended to fully describe a system or an experience. This was clear in Milgram and Kishino's conception as well: They supplemented their RV continuum with their less well-known taxonomy for characterizing mixed reality technology. In the next section, we revisit Milgram and Kishino's taxonomy and propose an updated version.

---

[3]https://www.microsoft.com/en-us/p/roboraid/9nblggh5fv3j
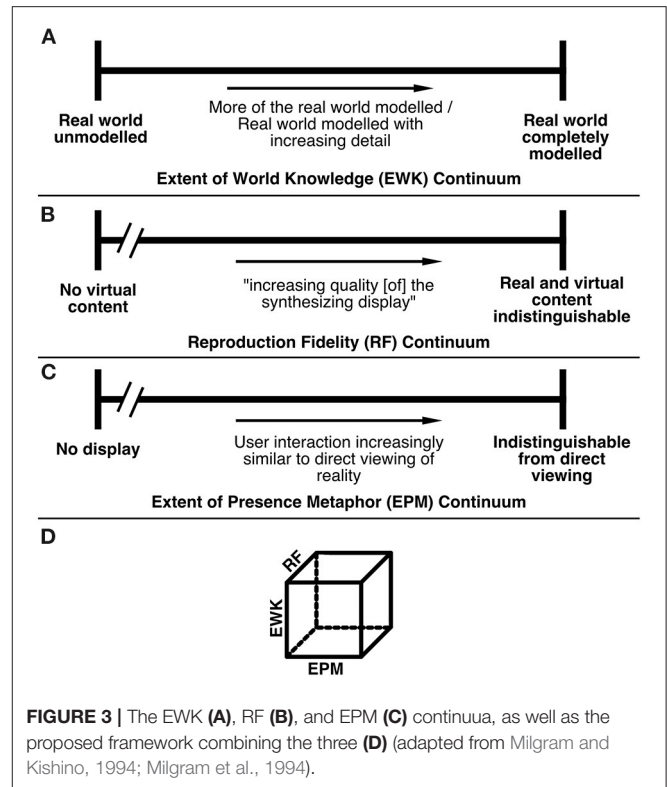
FIGURE 3 | The EWK **(A)**, RF **(B)**, and EPM **(C)** continuua, as well as the proposed framework combining the three **(D)** (adapted from Milgram and Kishino, 1994; Milgram et al., 1994).

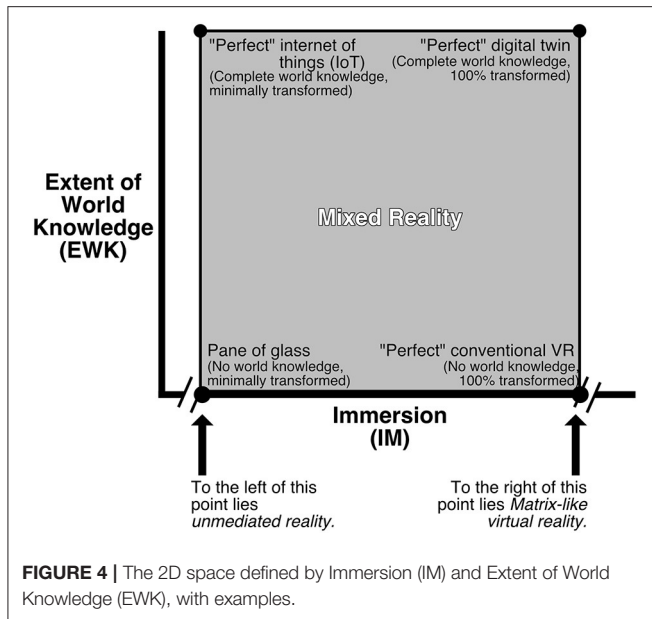# 4. A NEW TAXONOMY OF MR EXPERIENCES

4.1. Milgram and Kishino's Original Taxonomy of MR *Systems*

Milgram and Kishino's original paper included a three-dimensional taxonomy to characterize various mixed reality technologies. First, **extent of world knowledge (EWK)** described the level of modeling of the real world, and specifically the *where* (locations of objects) or *what* (identification of objects), included in the MR environment. Second, the **reproduction fidelity (RF)** of a technology described a display technology's capability of exactly reproducing the real world. Finally, the **extent of presence metaphor (EPM)** accounted for the level of world-conformal graphics and viewpoint of the person experiencing the MR environment (Essentially, the naturalness of the user's interaction with the display). These dimensions, and how Milgram and Kishino viewed their relationship, can be seen in **Figure 3**.

4.2. Our Proposed Taxonomy for MR *Experiences*

In the mid-1990s, both head-worn displays and computer hardware were generally bulky and, except for a few systems such as the Sony Glasstron and Virtual i-O i-glasses!, expensive. A typical research laboratory system minimally required a head-worn display, a high-performance workstation, and a tracking system. Total system cost could easily exceed 100,000

FIGURE 4 | The 2D space defined by Immersion (IM) and Extent of World Knowledge (EWK), with examples.

USD. Except for demonstration programs and a few games, most applications were custom developed, often on custom or customized hardware. Examples include Disney's virtual reality application *Aladdin* (Pausch et al., 1996), State et al.'s augmented reality system for ultrasound-guided needle biopsies (State et al., 1996), and Feiner et al.'s augmented reality application *Touring Machine* (Feiner et al., 1997).

In 2020, most mixed reality experiences can be implemented using off-the-shelf hardware solutions for components such as visual displays, processors, trackers, and user input devices. Even off-the-shelf display devices (including mobile phones or tablets used as AR displays, head-worn AR displays, and head-worn VR displays) often include tracking in addition to integrated processing. Therefore, except for custom systems, the main differentiator among MR experiences at near points on the continuum is no longer the mediating technology, but instead is the user's overall experience (This notion is echoed by Speicher, Hall, and Nebeling's interviewees: "[I]n the future, we might distinguish based on applications rather than technology," Speicher et al., 2019). In response to the major technology changes since the mid 1990s, we propose to modify and expand Milgram and Kishino's taxonomy in order to be able use it to categorize not only mixed reality technologies, but also, and importantly, mixed reality *experiences*.

Two of our proposed dimensions derive from Milgram and Kishino's original three. We adopt the **Extent of World Knowledge (EWK)** dimension directly, as we feel this captures a key component of augmented reality and augmented virtuality experiences—the extent to which the system is aware of its real world surroundings and can respond to changes in those surroundings. While, Milgram described EWK as a combination of *what* and *where* known objects are, modern sensing technologies, such as imagined in a pervasive *internet of things*, could provide access to much richer streams of
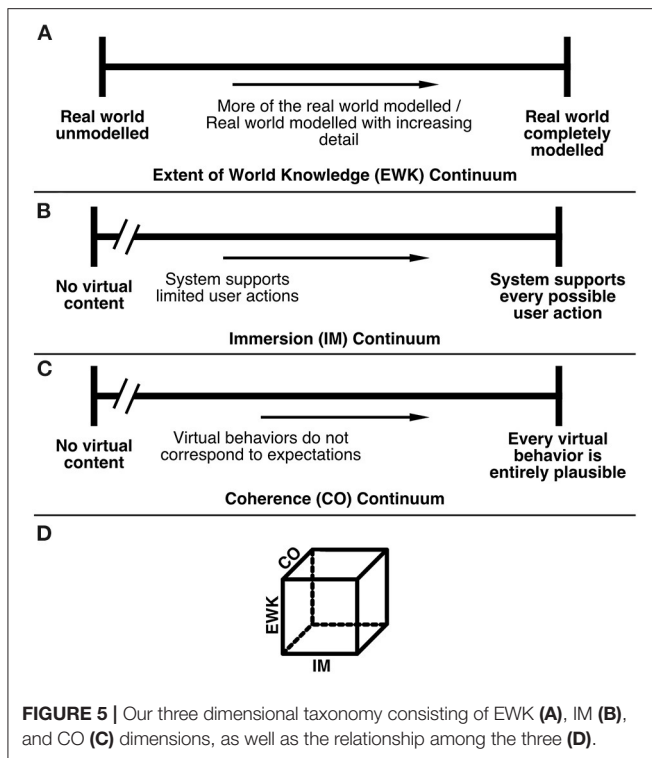
information about the real world environment. "Perfect" EWK would take advantage of these additional sensing capabilities wherever available and would extend to *how* things work, and *when* things might happen.

We propose to combine the Reproduction Fidelity (RF) and Extent of Presence Metaphor (EPM) dimensions into a single dimension, **Immersion (IM)**. We adopt the name following the definition of immersion favored by Slater et al. (Slater, 2004; Skarbez et al., 2017). That is, a system's immersion is the set of valid actions supported by that system. We choose to combine RF and EPM based on similarities between these dimensions, hinted at by **Figure 3** and remarked upon by Milgram and Kishino in the original paper: "...so too is the EPM axis in some sense not entirely orthogonal to the RF axis, since each dimension independently tends toward an extremum which ideally is indistinguishable from viewing reality directly" (Milgram and Kishino, 1994). Both the RF and EPM dimensions have a discontinuity at the minimum, as, when there is no display, the real world is perceived in an unmediated fashion, which, by definition, is indistinguishable from reality. We argue that this similarity in the two dimensions is not an accident; a system's immersion has the same behavior.

Furthermore, Slater has argued that immersion comprises two types of valid actions: *sensorimotor* valid actions and *effectual* valid actions. These are valid actions that result in changes to a user's perception of the environment and changes to the environment itself, respectively. We argue that both RF and EPM are actually part of sensorimotor valid actions. The original Milgram and Kishino paper does not account for the possibility of a system being interactive, beyond the choice of viewpoint. This limitation is also addressed by the inclusion of effectual valid actions in the combined dimension.

Already, with these two dimensions, IM and EWK, we can begin to productively characterize MR systems (**Figure 4**). External virtual reality systems generally have high immersion, with little or no world knowledge. Augmented reality systems, on the other hand, have low or medium immersion, but a higher level of world knowledge. A 2D mapping of these dimensions would show four extremes with IM along the x-axis and EWK on the y-axis. At the bottom left, a *pane of glass* represents no world knowledge or immersion at all. Perfect world knowledge with no immersion could be thought of as an *internet of things (IoT)* system, wherein the system knows the state of the real world, but does not itself display that state to a user. A fully immersive system with no world knowledge is the ideal of *conventional VR*, in which the virtual world is rendered exquisitely, but the system does not consider the real environment. The top right corner of the graph requires high IM and high EWK together, which results in a perfect *digital twin*, offering real-time tracking and rendering of the real world.

Populating this entire 2D IM-EWK space with examples is beyond the scope of this paper, as there are far too many to include. However, the upper-right portion of the space, near the extrema that we have labeled as a "perfect" digital twin, has not been as widely explored. One thread of research in this space is MR telepresence, in which the goal is to capture remote spaces and users and reproduce them elsewhere in real time. Recent

FIGURE 5 | Our three dimensional taxonomy consisting of EWK **(A)**, IM **(B)**, and CO **(C)** dimensions, as well as the relationship among the three **(D)**.

papers in this area include Kunert et al. (2018), Stotko et al. (2019b), and Stotko et al. (2019a). Another research thread is what we call world-aware environment generation, in which the goal is to create a virtual environment that possesses some of the same characteristics of the real environment—for example, areas that are not navigable in the real environment are also not navigable in the virtual environment. Recent research in this area includes Sra et al. (2016) and Cheng et al. (2019).

To describe the essence of a user's experience, and to address the third limitation identified in section 2, we must go beyond the elements of realism contained in EWK and IM to consider how sensory inputs create a unified experience–the coherence of the experience. We argue that this third dimension already exists in the literature. Various authors refer to it alternatively as *fidelity* (Alexander et al., 2005) or authenticity (Gilbert, 2017), but we prefer the term **Coherence (CO)**, following (Skarbez et al., 2017). Our resulting taxonomy is illustrated in **Figure 5**. While, EWK and IM can be thought of as describing what the system is *intended to do*, CO describes *how consistently that intention is conveyed to the user*. In the context of VR, coherence is primarily *internal*; that is, do virtual objects interact with one another and the user in predictable ways? In the context of AR, coherence is primarily *external*; that is, do virtual objects interact with real objects and the user in predictable ways? For example, in AR, objects meant to be fixed in the world would be externally incoherent if they float in space rather than sitting on a surface. In applications that are "truly" MR, both internal and external coherence are necessary for a satisfying and effective user experience.

4.3. Implications of Our Revised Taxonomy

In the remainder of this section, we will take it as given that our proposed taxonomy is appropriate—that it consists of three logically distinct, if not wholly orthogonal dimensions, and that these dimensions span a meaningful portion of the space of possible MR experiences. Each of the following subsections addresses a consequence which derives from this taxonomy.

4.3.1. Appropriate Constructs for Describing MR Experiences

We believe that Extent of World Knowledge, Immersion, and Coherence are objective characteristics of MR systems and that metrics could be identified for each of them. A variety of subjective constructs have been proposed to evaluate a user's experience of VR and AR systems, perhaps the most well-known of which is *presence*. Elsewhere in the literature, it has been argued that presence results from the combination of *Place Illusion (PI)*—otherwise known as *spatial presence*—and *Plausibility Illusion (Psi)*, with Place Illusion arising from the immersion of a system (Slater, 2009), and Plausibility Illusion from the coherence of a system (Skarbez, 2016). Even if one accepts that PI and Psi are appropriate constructs and that one has valid means to measure them—a difficult and contentious topic itself—neither of them contains any notion of extent of world knowledge. To our knowledge, there has been little or no research regarding EWK since Milgram et al.'s original papers on the topic. Certainly no constructs have been proposed that combine EWK with IM and/or CO. This is an area ripe for future research. We do not claim to have all the answers in this area, but in the interest of stimulating discussion, we propose the following model (**Figure 6**; constructs that have not yet been named or discussed in the literature are in *ITALICS*):
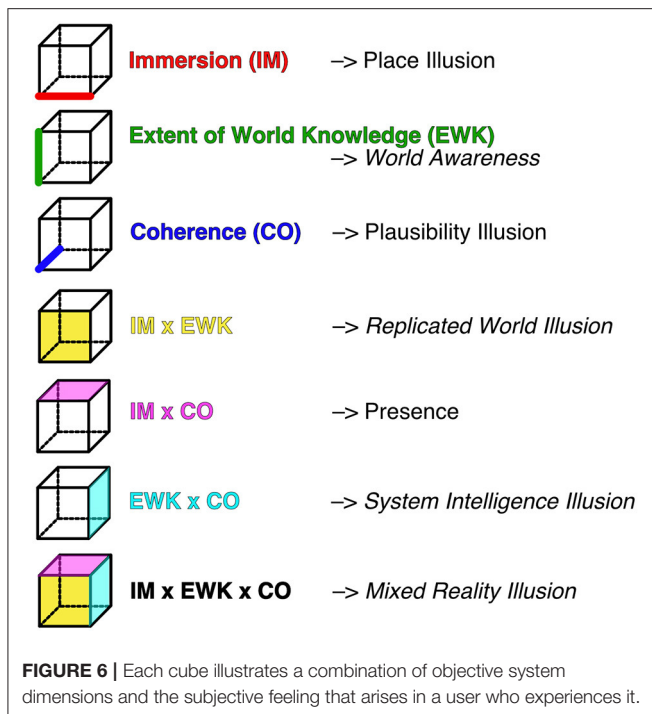
Some brief commentary on these new constructs:

- By *world awareness*, we mean a user's feeling that the system is aware of the physical world around them.
- By *replicated world illusion*, we mean a user's feeling that they are in a virtual copy of the real world (which may be analogous to the concept of *telepresence* as described in Steuer, 1992).
- By *system intelligence illusion*, we mean a user's feeling that the system itself is aware of its surroundings and uses that awareness intelligently; that is, in ways that do not violate coherence.
- By *mixed reality illusion*, we mean a user's feeling that they are in a place that blends real and virtual stimuli seamlessly and responds intelligently to user behavior.

These constructs are also useful in describing the discontinuities on the continua. We posit that having no immersion at all, or no computer-generated stimuli, means that MR is not possible. Thus, regardless of the quantity of EWK or coherence present, if immersion is non-existent, then so is the MR experience.

4.3.2. The Difficulty of Construct Measurement

Note that the objective nature of Extent of World Knowledge, Immersion, and Coherence doesn't mean that they are easy to measure. Far from it! It just means that they are characteristics of

FIGURE 6 | Each cube illustrates a combination of objective system dimensions and the subjective feeling that arises in a user who experiences it.

the *system*, not of the system's *user*. Speicher et al. also developed a conceptual framework for describing objective characteristics of MR, which include: number of environments, number of users, level of immersion, level of virtuality, degree of interaction, and input/output. While these dimensions can help describe a system, they do not provide guidance for evaluating users' experiences. However, by applying our 3D framework to the description of a given MR experience, we believe it may be possible to generalize recommendations for assessing the experience. Assigning specific values for a given system on any of the EWK, IM, or CO axes remains a substantial open research problem. For example, when considering the immersion continuum, should we place technology based on its immersion across all senses, or should each sense be measured in isolation (adding significant complexity as it would change our one-dimensional immersion continuum into six or seven dimensions)? Further, how do we measure how far along in the continuum a technology is? Categorizing MR experiences is difficult, but placing it in our 3D space can guide researchers and practitioners as to what measures (previously or currently used, or perhaps still to be developed) may be most appropriate for evaluating users' MR experiences.

4.3.3. Evaluating MR Experiences

In section 4.3.1, we proposed a set of constructs—some already well-represented in the literature, others described here for the first time—that could be used to describe a user's experience of an MR application. Under some circumstances, these constructs

may suffice for evaluation of such applications. In most cases, however, they will not. This is because, when evaluating an application, it is important to consider the intent of its creators. For a virtual reality game, it is more important whether it is *entertaining* than whether it is immersive. For a virtual reality stress induction protocol, it is more important that it is *stressful* and *controllable* than that it gives rise to presence. For an automotive AR head-up display, it is more important that it be *useful* and *safe* than for it to be highly world-aware. All of this is to say that evaluation is a different process than characterization; it requires different, and in many cases *application-specific* measures. However, accurate characterization helps us identify what measures may be most appropriate and valid for a given scenario.

5. CONCLUSION

Our intention with this article has not been to disparage Milgram and Kishino's work; on the contrary, we think it has admirably stood the test of time, and deserves to be recognized as one of the seminal papers in the field. That said, with the benefit of hindsight, there are some areas that we feel needed updating for the concepts to remain relevant in 2020 and beyond. To that end, we proposed a revised version of the reality-virtuality continuum based on the idea that virtual content is always ultimately situated in the real world, which has the consequence that conventional virtual reality should fall within the category of mixed reality. We argued for the continued relevance of a "big tent" definition of MR, and in fact, argued to make the tent bigger still by including all technology-mediated experiences under the term mixed reality. We presented a new taxonomy for describing MR experiences with the dimensions *extent of world knowledge, immersion, and coherence*. The new taxonomy was inspired by Milgram and Kishino's taxonomy which we feel has been underappreciated in comparison to the other contributions made by their original paper. Much as we were inspired by and are indebted to Milgram and Kishino's original work, we hope that this paper encourages further discussions and research in this area.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article.

AUTHOR CONTRIBUTIONS

RS, MS, and MW contributed to the ideation and development of this article. MS developed the initial draft, which was substantially revised and expanded by RS. RS, MS, and MW contributed to further revision and polishing of the final article. All authors contributed to the article and approved the submitted version.

REFERENCES

Alexander, A. L., Brunyé, T., Sidman, J., and Weil, S. A. (2005). "From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games," in *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (Orlando, FL: NTSA).

Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). "Haptic retargeting: dynamic repurposing of passive haptics for enhanced virtual reality experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16* (New York, NY: ACM), 1968–1979. doi: 10.1145/2858036.2858226

Azuma, R. T. (1997). A survey of augmented reality. *Presence Teleoperat Virtual Environ.* 6, 355–385. doi: 10.1162/pres.1997.6.4.355

Barba, E., MacIntyre, B., and Mynatt, E. D. (2012). Here we are! Where are we? Locating mixed reality in the age of the smartphone. *Proc. IEEE* 100, 929–936. doi: 10.1109/JPROC.2011.2182070

Cheng, L., Ofek, E., Holz, C., and Wilson, A. D. (2019). "Vroamer: generating on-the-fly VR experiences while walking inside large, unknown real-world building environments," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka), 359–366. doi: 10.1109/VR.2019.8798074

Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. *Pers. Technol.* 1, 208–217. doi: 10.1007/BF01682023

Flavián, C., Ibáñez Sánchez, S., and Orús, C. (2021). The influence of scent on virtual reality experiences: the role of aroma-content congruence. *J. Bus. Res.* 123, 289–301. doi: 10.1016/j.jbusres.2020.09.036

Gilbert, S. B. (2017). Perceived realism of virtual environments depends on authenticity. *Presence Teleoperat. Virtual Environ.* 25, 322–324. doi: 10.1162/PRES_a_00276

Heilig, M. L. (1962). *Sensorama Simulator.* U.S. Patent No 3,050,870. Washington, DC: U.S. Patent and Trademark Office.

Insko, B. E. (2001). *Passive haptics significantly enhances virtual environments* (Ph.D. thesis). University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.

Iwata, H., Yano, H., Uemura, T., and Moriya, T. (2004). "Food simulator: a haptic interface for biting," in *IEEE Virtual Reality 2004* (Chicago, IL: IEEE), 51–57. doi: 10.1109/VR.2004.1310055

Koleva, B., Benford, S., and Greenhalgh, C. (1999). *The Properties of Mixed Reality Boundaries.* Dordrecht: Springer. doi: 10.1007/978-94-011-4441-4_7

Kunert, C., Schwandt, T., and Broll, W. (2018). "Efficient point cloud rasterization for real time volumetric integration in mixed reality applications," in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Munich), 1–9. doi: 10.1109/ISMAR.2018.00023

Lindeman, R. W., and Noma, H. (2007). "A classification scheme for multi-sensory augmented reality," in *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology, VRST '07* (New York, NY: Association for Computing Machinery), 175–178. doi: 10.1145/1315184.1315216

Liu, S., and Manocha, D. (2020). Sound synthesis, propagation, and rendering: a survey. *arXiv [preprint]. arXiv:2011.05538.* Available online at: https://arxiv.org/abs/2011.05538

Mackay, W. E. (1998). "Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers," in *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '98* (New York, NY: Association for Computing Machinery), 13–21. doi: 10.1145/948496.948498

Mann, S., Furness, T., Yuan, Y., Iorio, J., and Wang, Z. (2018). All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv [preprint]. arXiv:1804.08386.* Available online at: https://arxiv.org/abs/1804.08386

Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inform. Syst.* 77, 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994). Augmented reality: a class of displays on the reality-virtuality continuum. *Proc. SPIE* 2351, 282–292. doi: 10.1117/12.197321

Miyashita, H. (2020). "Norimaki synthesizer: taste display using ion electrophoresis in five gels," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI), 1–6. doi: 10.1145/3334480.3382984

Normand, J.-M., Serviéres, M., and Moreau, G. (2012). "A new typology of augmented reality applications," in *Proceedings of the 3rd Augmented Human International Conference, AH '12* (New York, NY: Association for Computing Machinery). doi: 10.1145/2160125.2160143

Pausch, R., Snoddy, J., Taylor, R., Watson, S., and Haseltine, E. (1996). "Disney's aladdin: first steps toward storytelling in virtual reality," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96* (New York, NY: ACM), 193–203. doi: 10.1145/237170.237257

Salisbury, J. K., and Srinivasan, M. A. (1997). Phantom-based haptic interaction with virtual objects. *IEEE Comput. Graph. Appl.* 17, 6–10. doi: 10.1109/MCG.1997.1626171

Savioja, L., and Svensson, U. P. (2015). Overview of geometrical room acoustic modeling techniques. *J. Acoust. Soc. Am.* 138, 708–730. doi: 10.1121/1.4926438

Skarbez, R. (2016). *Plausibility illusion in virtual environments* (Ph.D. thesis). The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.

Skarbez, R., Brooks, F. P. Jr., and Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Comput. Surv.* 50, 96:1–96:39. doi: 10.1145/3134301

Slater, M. (2004). "A note on presence terminology," in *Presence Connect,* Vol. 3, 1–5. Available online at: http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm

Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Speicher, M., Hall, B. D., and Nebeling, M. (2019). "What is mixed reality?" in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19* (New York, NY: Association for Computing Machinery), 1–15. doi: 10.1145/3290605.3300767

Speiginer, G., and MacIntyre, B. (2018). "Rethinking reality: a layered model of reality for immersive systems," in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Munich), 328–332. doi: 10.1109/ISMAR-Adjunct.2018.00097

Sra, M., Garrido-Jurado, S., Schmandt, C., and Maes, P. (2016). "Procedurally generated virtual reality from 3D reconstructed physical space," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, VRST '16* (New York, NY: Association for Computing Machinery), 191–200. doi: 10.1145/2993369.2993372

State, A., Livingston, M. A., Garrett, W. F., Hirota, G., Whitton, M. C., Pisano, E. D., et al. (1996). "Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA), 439–446. doi: 10.1145/237170.237283

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *J. Commun.* 42, 73–93. doi: 10.1111/j.1460-2466.1992.tb00812.x

Stotko, P., Krumpen, S., Klein, R., and Weinmann, M. (2019a). "Towards scalable sharing of immersive live telepresence experiences beyond room-scale based on efficient real-time 3D reconstruction and streaming," in *CVPR Workshop on Computer Vision for Augmented and Virtual Reality* (Long Beach, CA), 3. Available online at: https://xr.cornell.edu/s/7_stotko2019cv4arvr.pdf

Stotko, P., Krumpen, S., Weinmann, M., and Klein, R. (2019b). "Efficient 3D reconstruction and streaming for group-scale multi-client live telepresence," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Beijing), 19–25. doi: 10.1109/ISMAR.2019.00018

Sutherland, I. (1965). "The ultimate display," in *Proceedings of the International Federation of Information Processing (IFIP)* (New York, NY), 506–508.

Yanagida, Y. (2012). "A survey of olfactory displays: making and delivering scents," in *Sensors, 2012* (Taipei: IEEE), 1–4. doi: 10.1109/ICSENS.2012.6411380