

Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces

Hans-Christian Jetter · Harald Reiterer · Florian Geyer

Received: 14 January 2013 / Accepted: 25 July 2013 / Published online: 18 October 2013
© Springer-Verlag London 2013

Abstract We introduce *Blended Interaction*, a new conceptual framework that helps to explain when users perceive user interfaces as “natural” or not. Based on recent findings from embodied cognition and cognitive linguistics, *Blended Interaction* provides a novel and more accurate description of the nature of human–computer interaction (HCI). In particular, it introduces the notion of *conceptual blends* to explain how users rely on familiar and real-world concepts whenever they learn to use new digital technologies. We apply *Blended Interaction* in the context of post-“Windows Icons Menu Pointer” interactive spaces. These spaces are ubiquitous computing environments for computer-supported collaboration of multiple users in a physical space or room, e.g., meeting rooms, design studios, or libraries, augmented with novel interactive technologies and digital computation, e.g., multi-touch walls, tabletops, and tablets. Ideally, in these spaces, the virtues of the familiar physical and social world are combined with that of the digital realm in a considered manner so that desired properties of each are preserved and a seemingly “natural” HCI is achieved. To support designers in this goal, we explain how the users’ conceptual systems use *blends* to tie together familiar concepts with the novel powers of digital computation. Furthermore, we introduce

four *domains of design* to structure the underlying problem and design space: *individual and social interaction, workflow, and physical environment*. We introduce our framework by discussing related work, e.g., metaphors, mental models, direct manipulation, image schemas, reality-based interaction, and illustrate *Blended Interaction* using design decisions we made in recent projects.

Keywords Blended Interaction · Post-WIMP · Natural user interfaces · Ubiquitous computing · Computer-supported cooperative work · Blends

1 Introduction

Interactive spaces are ubiquitous computing environments for a computer-supported collaboration that builds on and enhances the preexisting motor, spatial, social, and cognitive skills of groups of users. Typically, post-WIMP¹ computing devices, e.g., interactive walls, tabletops, tablets, smart phones, and post-WIMP interaction techniques, e.g., multi-touch, tangible, pen, or gestural input, are integrated into established non-digital work practices and work environments such as meeting rooms, design studios, or libraries. The goal is to achieve an ideally natural and unobtrusive computational support during collaborative activities such as presentation, ideation, analysis, or decision making. The virtues of physical and digital artifacts are combined in a considered manner so that desired properties of each are preserved, and a seemingly “natural” human–computer interaction (HCI) is achieved. In conclusion, interactive spaces purposefully *blend* the power of

H.-C. Jetter (✉)
Intel ICRI Cities, University College London, Gower Street,
London WC1E 6BT, UK
e-mail: h.jetter@ucl.ac.uk

H. Reiterer · F. Geyer
Human–Computer Interaction Group, University of Konstanz,
Universitaetsstrasse 10, Box D73, 78457 Konstanz, Germany
e-mail: Harald.Reiterer@uni-konstanz.de

F. Geyer
e-mail: Florian.Geyer@uni-konstanz.de

¹ Post-WIMP = Post-“Windows Icons Menus Pointer”.

digital computing with natural work practices and collaboration styles.

The successful design of such interactive spaces poses a wide range of challenges that are concerned with many technological problems, but in particular with yet unanswered research questions of HCI and interaction design. The crucial question is how to blend computational power with non-digital practices of natural collaboration without disrupting these practices. As we all have experienced in our daily life, computers and their user interfaces (UIs) tend to introduce unnatural, complex, seemingly unpredictable, or even “obnoxious” [1] behavior into our workplaces. For example, Oulasvirta [2] observed users “do” present-day ubicomp and characterizes it as “a multi-layered agglomeration of connections and data, distributed physically and digitally, and operating under no recognizable guiding principles.” Using computing technology still requires special skill sets and often, instead of supporting users in taking out their actual work, computing devices distract users toward the secondary task of using, managing, and configuring them. Nonetheless, computers can of course provide powerful support during collaboration, for example when creating, editing, publishing, and searching knowledge artifacts. If designed appropriately, they enormously increase the users’ productivity. Therefore, an often-cited goal of HCI is to create an “invisible” or “calm” ubiquitous computing in the sense of Weiser [3] that provides us with the great powers of digital computation in an unobtrusive manner, so that we are freed to use them without thinking and “mental gymnastics” and to focus beyond computers on new goals.

Unfortunately, in spite of the recent years’ rapid progress in post-WIMP technologies and many examples of successful interaction design for single post-WIMP applications or systems, we still lack a deeper understanding of why some interfaces are successfully becoming “invisible” and others fail. This situation has not improved since over a decade, when Whittaker et al. [4] already criticized that we do not have a good understanding of and lack theories about why certain core interface techniques are successful. For example, HCI researchers tried to establish predictive models based on cognitive architectures such as GOMS or ACT-R [5] but both are only very rarely used in design practice. The reason for this is that they naturally work at an extremely detailed level that involves laborious modeling even for the simplest user tasks [6]. Even when different post-WIMP designs would be modeled at this level, GOMS or ACT-R only expose currently unresolved psychological issues to the modeler for resolution instead of helping to solve them [6], for example, the disputed benefits of metaphors, direct manipulation, multi-touch, or tangible user interfaces (TUIs) [7–10]. Therefore, it is fair to say that HCI researchers still do not understand why

some post-WIMP designs are perceived as “natural” or “intuitive,” while others are not because there is no theory, model, or framework about the cognitive processes that let us perceive UIs this way or the other.

It is a legitimate question to ask why we should bother with theory at all. After all, unlike predictive models, usability testing and user studies have proven to be valuable tools during design practice and are successfully used for iteratively improving interaction designs. Nonetheless, we believe that “you cannot test your way to quality” [11]: A better understanding of cognitive processes, even if it is applied only indirectly and informally, remains desirable, because it enables entering the iterative design process with better designs and facilitates the interpretation of the empirical observations made therein. In our view, today’s research in HCI tends to focus too strongly on creating novel designs and technologies and only formulates design recommendations for selected problems or application domains. We believe that HCI as a scientific discipline cannot restrict itself to solely creating this patchy body of theory on a case-by-case basis that is comprised of individual cases of design and technology. It must also strive for novel theoretical frameworks that are connecting these dots. Of course, this call for a “unified theory” in HCI is not new, and we do not believe that a “theory of everything” in HCI is possible. However, we believe that the aforementioned challenge of understanding and designing a “natural” post-WIMP HCI with ubiquitous computing that is based on blending the real with the digital world is a realistic goal. In this article, we therefore provide our theoretical framework of *Blended Interaction* that is based on consolidating and extending previous and related theory from HCI, cognitive science, and cognitive linguistics, and we illustrate this framework using examples of successful post-WIMP designs from our work.

2 Why are computers hard to use?

Our starting point in this article is an admittedly obvious but important question: Why are computers hard to use? Although this question might appear overly simplistic at first sight, it is by no means trivial to answer. A serious answer must involve disciplines such as cognitive science and psychology, since we can learn from them how we as humans understand, adopt, and use new tools. More concretely, contemporary cognitive science reveals a fundamental pattern underlying human cognition that is always applied when we learn new concepts, regardless of whether they are from the non-digital or digital realm. This fundamental pattern in human cognition provides us with the theoretical foundation for *Blended Interaction* and is introduced in the following.

2.1 Embodied cognition

Contemporary views from cognitive science and cognitive linguistics such as *experiential* [12: p. 19] or *embodied cognition* [13] teach us that our abilities for and experiences of interacting with our natural environment, e.g., physical interaction with objects or social interaction with other individuals, are also defining the ways in which we can think and reason about the world. Whether we like it or not, our human body and its primate ancestry, our sensory system, our perception of space and time, our abilities for physical and social interaction, and similar evolutionary and developmental influences ultimately define and restrict what our mind can learn, understand, and make use of. “We divide the world up into entities at human scale so that we can manipulate them in human lives [...]. Frogs and bats, for example, divide the world up in ways quite different from our own” [14: p. 8]. Imaz and Benyon highlight the resonances of this fundamental shift in cognitive science with Wittgenstein’s famous pronouncement that even “if a lion could talk, we could not understand him.” They write: “The experience of a lion is so different from that of a human that the lion would have vastly different concepts” [12: p. 19]. According to Dourish [15: p.vii], this new approach in cognitive science and philosophical psychology abandons the idea of disembodied rationality and, instead, practical action and everyday experience replaces abstract reasoning and objective meaning.

As a consequence, even the things that we typically consider as highly abstract and intellectual such as language use, philosophical or mathematical concepts, or expertise in crafts or sciences are also crucially shaped by our bodies and brains, especially by our sensorimotor system [16: p. 22]. According to linguists Lakoff and Johnson [17: p. 3], the same concepts that govern our everyday functioning, e.g., what we perceive, how we get around in the world, and how we relate to other people, are also governing our intellect, thinking, and language. In this view, even highly complex concepts such as “guilt” or “hyperlink” are part of the same conceptual system that also contains basic-level sensorimotor concepts such as “up” and “down,” “large” and “small,” or “heavy” and “light.” Highly complex concepts in our conceptual system are generated from less complex or basic-level concepts by metaphorically integrating or *blending* them. This process of *blending* or *conceptual integration* (see Fauconnier and Turner [14]) enables us to innovate and to create new, more complex concepts from existing concepts as input. The resulting output concept has a new emergent structure that is not available from the inputs alone. Over time this creates a vast network of conceptual integrations that connects high-level complex concepts via many intermediate steps

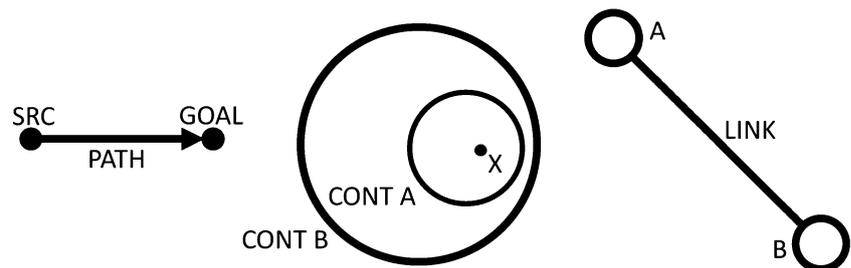
to our most basic low-level bodily, spatial, or social experiences. Although we use this network in almost every moment of thinking, we are not consciously aware of it. Integration typically goes entirely unnoticed since it works fast in the backstage of cognition [14]: p. 6]. “Metaphoric thinking, regarded in the commonsense view as a special instrument of art and rhetoric, operates at every level of cognition and shows uniform structural and dynamic principles, regardless of whether it is spectacular and noticeable or conventional and unremarkable” [14: p. 17]. In summary, contemporary *embodied cognition* teaches us that “The mind is inherently embodied. Thought is mostly unconscious. Abstract concepts are largely metaphorical. These are the three major findings of cognitive science” [16: p. 3].

2.2 Natural versus digital

So how do these findings help us to answer our question why computers are hard to use? In our interpretation of embodied cognition, a concept is hard to learn, understand, and apply when it involves too many steps of conceptual integrations in our conceptual system. The more a concept is distant from the already existing concepts, the greater is the necessary number of steps and the intellectual effort for integrating it. The closer a concept is to existing concepts, in particular to the core basic-level bodily, spatial, or social concepts that most of us share since our childhood, the easier it is to integrate and apply. This is not only true for the concepts that we use to think, reason, and act in the non-digital world, but it also shapes the ways in which we are able or unable to conceptualize, understand, use, and adopt digital technology and its interactive functionality. As a consequence, the concepts of a computer UI that are closer to our shared everyday bodily, spatial, or social experiences are easier for us to learn, understand, and apply, since we have to commit fewer cognitive resources.

In their seminal cognitive account of direct manipulation UIs from the 1980s [9], Hutchins et al. assume that a feeling of “natural” and “direct” interaction results from a commitment of fewer cognitive resources to bridge the gulfs of execution and evaluation. Jacob et al. [18] *reality-based interaction* transfers this assumption into the post-WIMP world. Similarly, we believe that the desirable feeling of a direct and “natural” interaction with a post-WIMP UI can be achieved by designing UIs whose concepts are close to a user’s already existing concepts and basic-level experiences. This is not only a question of simple visual familiarity, i.e., quickly recognizing which part of a UI (e.g., icons, symbols) represents which already familiar entity or activity, but also of understanding the more abstract mechanisms of how the parts interact with or constrain each other, e.g., whether they are logically or

Fig. 1 Examples for *image schemas*: SOURCE–PATH–GOAL (=“something moves from SRC along PATH to GOAL”), CONTAINER (=“if X is in A and A is in B then X is in B”), LINK (“A is linked to B”) [12]



physically connected, if one contains the other, if objects remain to exist after becoming invisible, or if actions are reversible or not. Furthermore, this is also true for the different concepts of navigation and manipulation in a UI, e.g., direct manipulation of a view or object by physically touching, grabbing, and moving it or indirect manipulation using commands or formal languages.

This view resonates with the work of Hurtienne and Israel [19] who design TUIs in accordance with the user’s basic-level knowledge on the sensorimotor level. We learn about basic-level concepts on the sensorimotor level (e.g., movement, containment, and linkage) very early in childhood. This knowledge is captured in our conceptual system as *image schemas* (Fig. 1). Image schemas are embodied prelinguistic structures of experience that cognitive linguists revealed by analyzing the use of metaphors in different languages. To this day, there are about 30–40 image schemas distinguished in literature that seem to exist across different populations, cultures, and languages. They range from “basic schemas” (e.g., SUBSTANCE, OBJECT), “containment” (e.g., IN–OUT, CONTENT, and FULL–EMPTY), “spatial motion” (e.g., PATH and SOURCE–PATH–GOAL) to “attributes” (e.g., HEAVY–LIGHT, WARM–COLD, and STRONG–WEAK).

Image schemas are so basic concepts that they exist on the sensorimotor level beneath our conscious awareness, but they enable us to structure more abstract concepts [19]. They function as the basic building blocks of thought, and we all share them and constantly use them beneath consciousness [20]. By metaphorically combining and transferring them to different domains, we generate and understand new and more complex concepts. Basic-level core concepts such as the SOURCE–PATH–GOAL or CONTAINER image schemas are among the first concepts that we learn in our life, for example, while we are interacting with baby toys and thus learn about physical cause-and-effect phenomena. We use image schemas to integrate these new experiences into our conceptual system. Their existence can explain why toddlers are already capable of selecting and moving virtual objects on the touchscreens of tablets: Image schemas help to transfer and apply the sensorimotor experiences from physical space, e.g., when

moving toys around, to the flat screens of the digital realm. Later in child development, image schemas are used as building blocks for growingly complex concepts that are created by conceptual integration (e.g., object permanence, new notions of location and containment, and simple math), which themselves serve later as the basis for high-level complex concepts that we would consider a part of culture and expertise.

However, the experiences that we make in our non-digital physical or social environment and that we integrate in our conceptual systems are usually very different from those that we make with the digital world of computation. For example, unlike real-world objects, a digital object can have a multitude of copies and instances that exist at different points in time and space. Often they are not physically contained in the devices that we use to access them, but they are located on something called a “server” or a “cloud” on the other side of the planet. Furthermore, digital objects and their state in main memory are non-persistent by default, and it needs additional computational infrastructure to let them imitate one of the most fundamental principles of the physical world, namely that “objects do not change their state or location, if there is no force applied to them,” a rough equivalent to Newton’s first law of motion. Even trained computer scientists often experience this mismatch between the non-digital and the digital realm, leading to a perception of computers as something unnatural, complex, unpredictable, or even magical. While the digital world “empowers users with creative and magical potential” [21], it follows its very own logic that often stands in stark contrast to our experience of the fundamental principles of our natural physical and social world.

Appropriately blending these opposing worlds is difficult for designers of UIs. The overall goal is to design interactions that, on the one hand, build upon our knowledge of the sensorimotor level and our familiarity with our non-digital physical and social environment and, on the other hand, fully leverage the virtually infinite possibilities, computational power, and “magic” of the digital world. To achieve this, HCI researchers and practitioners have tried to use *analogies* in UIs to convey digital functionality in forms that at least visually look

like non-digital objects or tools. For example, the desktop interface uses a *user interface metaphor*² to hide the computer's complexity and presenting functionality in seemingly familiar non-digital objects, e.g., office desks, folders, or paper documents. However, metaphors are also often criticized and many argue for less literal but more abstract and “metaphor-free” approaches [10]: For example, Nardi and Zamer [22] suggest using *visual formalisms* instead of metaphors. They are computational building blocks, e.g., interactive tables, graphs, plots, panels, or maps that present application semantics in a clear, precise way. Bederson et al. [23] suggest using *informational physics* instead of mimicking the mechanisms of earlier media to more fully exploit radical new computer-based mechanisms. For example, they introduce zoomable UIs with the analogy of a “miraculous new material that is stretchable like rubber but continues to display a crisp computer image, no matter what the sheet's size.” More recently, in *reality-based interaction*, Jacob et al. [18, 24] do not refer specifically to any of these approaches, but identify four generic *themes of reality* from the physical, non-digital world that can be used to build on users' preexisting knowledge and to make computer interaction more like interacting with the real, non-digital world.

In this paper, we do not argue for or against any of these approaches. We believe that all of them are appropriate design tools depending on the users' and the designers' background, skills, context, and task. Some of them might start with an imprecise or false notion of human cognition—for example, in the light of Lakoff and Johnson [16, 17], there is nothing like a “metaphor-free” UI as there is also no “metaphor-free” thinking. Nevertheless, they have been successfully applied in the past, so that there must be some value to them. We are interested in the commonalities between the different approaches and want to understand what the underlying cognitive mechanisms are that make them work or fail. We are seeking for an overarching model that lets us understand when analogies are understood by users and when they are not. For this reason, we first have a closer look at how embodied cognition and cognitive linguistics model the human conceptual system and how humans create and use analogies with *conceptual integrations* or *blends*.

3 The way we think: conceptual integration

As mentioned, new concepts are based on a metaphorical combination of existing concepts. But “how can two ideas

² Blackwell provides an extensive review on the role and history of user interface metaphors as a design tool in HCI in [13].

be merged to produce a new structure, which shows the influence of both ancestor-ideas without being a mere ‘cut-and-paste’ combination?” [25]. According to linguists Fauconnier and Turner, whose work builds upon that of Lakoff and Johnson, our ancestors developed this skill around fifty-thousand years ago. “They acquired a modern human imagination, which gave them the ability to invent new concepts and to assemble new and dynamic patterns” [14: p. v]. The great mental capacity behind this is called *conceptual integration* or *conceptual blending*. Conceptual integration creates a new concept from two existing input concepts. The resulting output concept is called a *blend* and has an emergent new structure that is not available from the inputs. One of three introductory examples that linguists Fauconnier and Turner use to explain conceptual integration and the resulting blend is taken from interaction design: the desktop UI. This is helpful for us in two ways. First, we can explain conceptual integration to the reader using a familiar example from HCI. Second, the example shows that Fauconnier and Turner regard understanding and using a UI as a result of conceptual blending. Their blending theory therefore seems particularly appropriate for HCI and as a basis for *Blended Interaction*.

In the view of Fauconnier and Turner, the user of the desktop manipulates an integrated structure that derives some of its properties from different inputs—office work, commands, and menus. “But however much the interface takes from the inputs, it has considerable emergent structure of its own: Pointing and clicking buttons is not at all part of traditional office work or choosing from lists of words on paper; having little two-dimensional squares disappear under other little squares is not part of giving commands or of putting sheets of paper into folders; and dragging icons with the mouse is not part of moving objects on a desktop, ordering a meal, giving standard symbolic commands, or, *a fortiori*, using the machine language” [14: p. 23]. The user manipulates the desktop UI³ “not by means of an elaborate conscious analogy but, rather, as an integrated form with its own coherent structure and properties. From an “objective” point of view, this activity is totally novel—it shares very few physical characteristics with moving real folders, and it is novel even for the traditional computer user who has issued commands exclusively from a keyboard rather than from a

³ Please note that, unlike HCI literature, Fauconnier and Turner do not refer to the desktop UI as the “desktop metaphor”. In their terminology, the desktop UI is a blend and not a metaphor, as is also discussed by Imaz and Benyon: “When we speak of the desktop metaphor now, we are really referring to a large blend [...]. It is usually considered to be a metaphor because most of the traditional functionalities of ordinary work have been maintained as expressions in interface tasks [...]. But when observed in detail, it is evident that we are dealing with a blend rather than a metaphor—the blend being based on the metaphor” [5: p. 52].

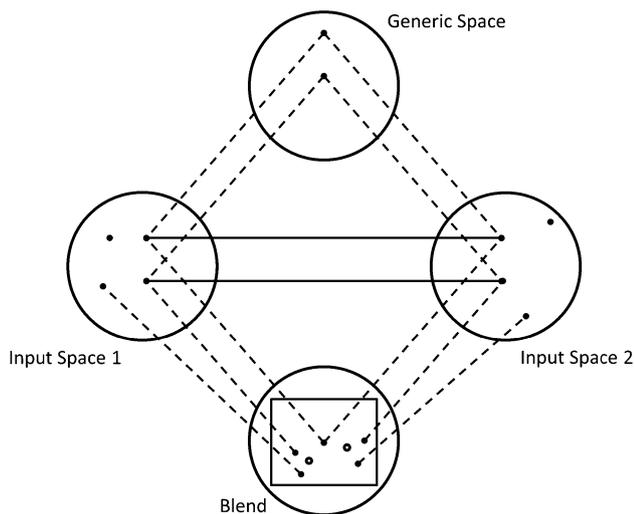


Fig. 2 A basic diagram of a blend. Adapted from [14: p. 46]

mouse. Yet the whole point of the desktop interface is that the integrated activity is immediately accessible and congenial” [14: p. 24].

In comparison with the traditional idea of a UI metaphor, the new notion of a blend is liberating for the designer. Since the blend has an emergent structure of its own, it is not necessary that a UI is a literal or precise simulation that correctly represents all aspects of a familiar concept. Blends explain how a UI that shares only selected aspects with its users’ reality can still be understood and adopted, even though some of its behaviors might appear very unfamiliar or unnatural. Thanks to the users’ ability of conceptual integration, designers are not bound to design an UI as realistic as possible, and they can avoid cases of “silly simulation” [26]. Instead, they can focus on a consistent, coherent, and congenial overall structure. However, conceptual integration must not serve as an excuse for using just any random combination of realistic and unrealistic elements for design, since our ability for conceptual integration has its limits and not just any discordant combination can be projected to the blend: “Some discordant structure is irrelevant because it has no bad consequences—for example, the trashcan and the folders both sit on the desktop—but other discordant structure is objectionable. Dragging the icon for a floppy disk to the trash as a command to eject the disk from the drive is notoriously disturbing to users” [14: p. 24].

In their work, Fauconnier and Turner, and Imaz and Benyon [12: p. 44; 14: p. 46] provide a more formal description of conceptual integration and a visual illustration of it (see Fig. 2): A *blend* is the output of conceptual integration and is based on two *input spaces* from different domains. For example, *input space 1* could be “physical office folder” and *input space 2* “computer file system.”

There is partial matching between the two *input spaces*, e.g., some elements of *input space 1* and some of *input space 2* are matched (e.g., “physical folders and folders of a computer file system are both containers,” “physical folders and folders of a computer file system can both be labeled”). The solid lines in Fig. 2 represent counterpart connections produced by matching.

These connections become possible by means of a *generic space*. The *generic space* contains abstract information that is common to both the inputs. Imaz and Benyon describe the *generic space* as some more abstract structure and organization shared by both *input spaces* that defines the core cross-space mapping between them [12: p. 44]. A given element in the *generic space* (e.g., the concepts “container” or “label”) maps onto paired counterparts in the two *input spaces*.

The *blend* (e.g., the “folder” of the desktop UI) is based on selective projection by the users who are “running the blend” in their mind. Thereby, not all the structure from the inputs is projected to the *blend*. As illustrated in the figure, some elements from the *input spaces* are not projected at all (e.g., “a physical folder can only contain a finite number of items,” and “before putting an item in a physical folder, holes must be created in them using a hole punch”). Other elements are projected to the blend although they only exist in one input space (e.g., “a physical folder can be opened or closed,” and “folders of a computer file system can contain other folders”). This results in the blend’s *emergent structure* that is more than a mere “cut-and-paste” combination. Its structure is not copied there directly from any input.

4 Using blends in human–computer interaction

When we encounter metaphors in language, e.g., “THIS SURGEON IS A BUTCHER” [12: p. 48], or in interaction design, e.g., “THE OPERATING SYSTEM IS AN OFFICE DESKTOP” [12: p. 52], a cognitive process of conceptual blending is triggered in our mind that generates an emergent structure from the provided input spaces, the generic space, and our preexisting knowledge of their elements. This is not a deterministic process with a clearly defined, guaranteed outcome, but the resulting emergent structure depends on the existence of similar mental spaces in our conceptual systems and how we match and selectively project them. Only if these spaces exist in our conceptual systems and are semantically close to that of the author or designer of the blend, we will generate an emergent structure that is similar to the one that the author or designer has intended.

This resonates with Norman’s notion of *mental* or *conceptual models* in UI design [27: p. 16]: Designers use their

own mental *design model* of the system to construct a *system image*, i.e., the system’s physical structure such as its UI. The *system image* is all that the system’s users can perceive, and by interacting with it, they start to develop their own mental *user’s model* that is ideally, but not necessarily, similar to the *design model*. “The designer expects the user’s model to be identical to the design model. But the designer doesn’t talk directly with the user—all communication takes place through the system image” [27: p. 16]. Neither Norman’s *mental models* nor the *blends* of *Blended Interaction* can be transferred directly from the designer to the user. Instead, they must be conveyed via the system’s design or image and are the results of a non-deterministic generative process that is based on the users’ knowledge and their interaction with the system.

4.1 Blends versus real-world metaphors

In today’s globalized markets of digital consumer products, it is not certain that designers and users share similar mental spaces, in particular if the design is based on elaborate conscious analogies. For example, younger users might not be familiar with the physical objects and tools from office environments (e.g., folders and file cabinets) when they are first confronted with a desktop interface. They also do not recognize the floppy disk icon on the “save” button in a word-processing application as a real-world object, because they have probably never seen or used such a disk in real life. For them, the benefit of the desktop UI lies mainly in the aforementioned “integrated form with its own coherent structure and properties” [14: p. 24] and not in the use of supposedly familiar physical objects or tools from office work. This resonates with the work of other HCI researchers who believe that the desktop interface’s benefits derive almost completely from direct manipulation rather than from metaphor [7]. Some would even argue that the desktop interface works despite and not because of real-world metaphors [10].

For the designers of a system, it is therefore important to match input spaces via generic spaces that are likely to be shared with them by the users and that lead to congenial blends with a coherent structure and properties. In our view, particularly good candidates are concepts on the sensorimotor level that we learn early in our lives—often without being aware of it—and that we all share because of their universal nature, e.g., the aforementioned *image schemas* [19] or also Norman’s interpretation of *affordances* [27: p. 219]. In a similar fashion, the four *themes of reality* of Jacob et al.’s *reality-based interaction* [18] are good candidates (Fig. 3):

1. The first theme of *naïve physics* is “the informal human perception of basic physical principles, or in

other words, common sense knowledge about the physical world. This includes concepts such as gravity, friction, velocity, the persistence of objects, and relative scale” [18]. Jacob et al. convincingly illustrate how physical principles such as inertia and springiness are used in the iPhone’s UI to facilitate many tasks, e.g., fast flicking through the contact list.

2. The second theme of *body awareness and skills* enables designers to use coordinated movements as user input, e.g., two-handed or whole-body interaction instead of simple button or key presses. Writing with a pen is a good example of a highly developed body skill and coordinated movement. It comes about through the coordinated use of pen, paper, and ink and involves the dominant hand for moving the pen while the non-dominant hand is used to orient the page or device appropriately [15: p. 51].
3. The third theme of *environment awareness and skills*, i.e., our natural skills to navigate and alter our spatial environment, can be used to provide space as a cognitive resource, e.g., for sorting or clustering objects during sense making [28, 29] or for using spatial configurations to organize and memorize objects in space and scale [8].
4. Finally, “people are generally aware of others in their environment and have skills for interacting with them” [18]. This theme of *social awareness and skills* enables us, for example, to collaborate in a shared physical space, e.g., when sitting around an ordinary table [30] or an interactive tabletop [31], without disrupting work by invading other people’s personal territories, taking away their physical tools, or interrupting their conversations.

The point we want to emphasize here is that interaction designers should consider using and blending the vast amount of concepts (e.g., image schemas and themes of reality) that we as humans share due to the similarities of our bodies, our early upbringing, and our sensorimotor experiences of the world before resorting to elaborate conscious analogies such as the desktop metaphor. Sometimes “having little two-dimensional squares disappear under other little squares” [14: p. 23] is sufficient to convey the necessary notions of space and containment in an UI simply by using image schemas such as CONTAINMENT, IN–OUT, CONTENT, FULL–EMPTY, FRONT–BACK, and SOURCE–PATH–GOAL. Domain-specific or culture-specific analogies such as “THE OPERATING SYSTEM IS AN OFFICE DESKTOP” that are based on the recognition of real-world objects such as “folders” or “windows” are often not necessary or even disrupting. Like Dourish has formulated for embodied interaction, a key

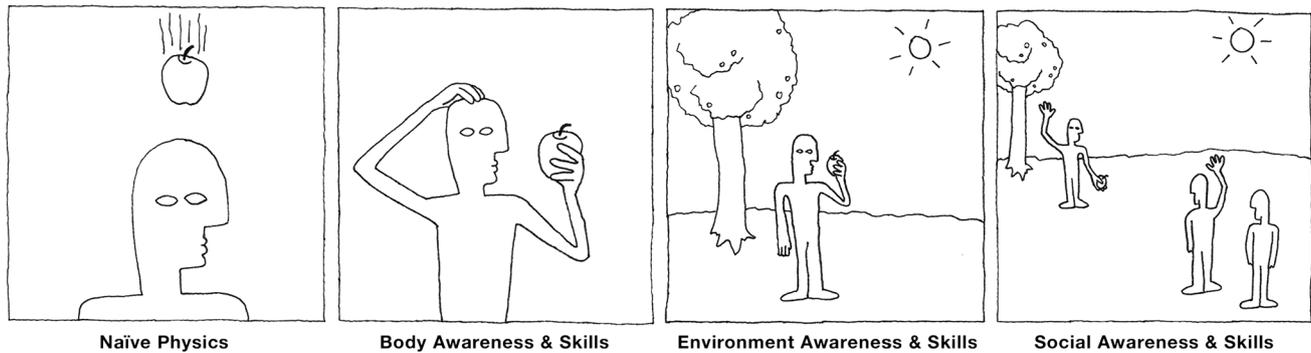


Fig. 3 The four *themes of reality* of Jacob et al. [18]

principle of *Blended Interaction* is as follows: “Instead of drawing on artifacts in the everyday world, it draws on *the way the everyday world works* or, perhaps more accurately, *the ways we experience the everyday world*” [15: p. 17].

4.2 What is still “real” in a digital world?

So far, we strongly focused *Blended Interaction* on building on users’ preexisting knowledge of the everyday, non-digital world, and our sensorimotor experiences of it. This seems to imply that we clearly distinguish between the non-digital, physical “reality” that is easy to understand because we are all familiar with it, and the “unreal” and “magic” world of computing whose disembodied concepts can only be understood after committing far more cognitive and intellectual resources.

However desirable such a clean separation may be, it would lead to a too simplistic view: We believe that we cannot consider our “reality” as free from digital influences anymore. Even though embodied cognition emphasizes how strongly our bodies and brains shape our cognition and how they are a result of millions of years of evolution, it also considers our cognitive abilities as socio-culturally determined by our prior experiences. Today, this inevitably includes our experiences with the computers and digital technologies that surround us. We believe that as effortlessly as users apply familiar concepts from the non-digital world (e.g., force, friction, and persistence of objects), they can also apply some of the concepts that they have learned from familiar digital technology (e.g., tap an object to issue an action, two-finger zooming and pinching, and undo/redo). Our conceptual systems internalize real-world and digital concepts likewise, and thus, our “reality” is not only based on the non-digital, physical world but is also constantly changed and influenced by popular digital products. When interacting with new technology, users apply all their preexisting knowledge, regardless of whether it stems from the physical or digital realm. Given the many examples of how humans have adopted digital

technologies and turned them into a fundamental part of their lives, excluding already well-established digital concepts from designing *Blended Interaction* would not reflect how users approach new UIs.

We have frequently witnessed this when watching users of (multi-) touch smart phones or tablets in museums or public spaces: Often these users have developed the expectation that all screens including that of exhibits or kiosk systems are touch-sensitive and thus they frequently try to change screen content by touch or using their fingers for sliding or pinching gestures—even when the exhibits are not touch-enabled at all. This can also be witnessed in interactive spaces where users are confronted with (and are often annoyed by) multiple displays of which some are touch-sensitive while others are not [32]. Within the few years of its existence, users have deeply internalized the concept of interaction by (multi-) touch as if it was a feature of the physical world, although there is not a single non-digital artifact in the world that changes its content or size by touching, pinching, or stretching it. Today’s toddlers and children are inevitably exposed to (multi-)touch technology, either by using it or watching its use. For them, touching or pinching screen content will be as “real” as it is for us today to turn a doorknob for opening a door or to push the button next to it for ringing the doorbell.

Viewing HCI through the lens of conceptual integration and *Blended Interaction* explains this deep internalization of supposedly “unnatural” concepts: For example, present-day users are very familiar with the concept of a digital “folder.” Although originally rooted in an analogy to a real-world office object, today, the concept of a “folder” has developed into a blend with an emergent structure of its own. In this structure, there are only very few of the original real-world elements left, so that the digital folder can hardly be considered a real-world metaphor anymore. Except for its name, it is largely disconnected from its real-world origin. Nevertheless, most users—and in particular young users that grew up with digital technology—have no problems with applying this concept across different

systems, e.g., PCs, laptops, digital cameras, smart phones and digital TVs and are not aware of its non-digital physical origin from office work. A further example is the undo/redo functionality known from office applications. The ability to undo/redo changes in the digital realm has become part of the users' reality, even though it is entirely impossible to achieve in the non-digital world since it contradicts fundamental laws of physics. Undo/redo has become such a deeply integrated part of the users' conceptual systems that it is generally preferred to manually recreating a previous state, even in those cases where this could be done almost effortlessly with a few interactions in a direct manipulation or tangible UI. The "magical" ability to jump back and forth in time to different states of a system or object has been widely adopted by users and is generally preferred to manipulation. This is a further example of how something entirely "unnatural" has become a part of our everyday experience and reality.

Frequently used blends from the digital world such as the "folder" or undo/redo have already become a part of the conceptual system of a large part of the user population and thus have become well established in their everyday world or reality. This enables designers to build new blends based on these already established blends, e.g., a "shared folder" that is shared via a server and thus exists at multiple locations at the same time or an undo/redo functionality for physical objects or TUIs based on actuators or reconfigurable materials. As we have already discussed before, blends that are closer to basic-level bodily, spatial, or social experiences are easier to understand. However, this does not mean that more distant blends (or blends of blends) are impossible to learn and apply. Command line interfaces and programming languages are probably the examples of the greatest distance between a UI and basic-level sensorimotor experiences. Nevertheless, they work for many users after sufficient training and learning, similar to the way students are able to learn algebra or calculus over their years in school although both seem not to have much in common with our physical or social world.

In conclusion, *Blended Interaction* explains the benefit of UI concepts that are close to our basic-level sensorimotor experiences of the non-digital, physical, or social reality without restricting the design space to exclusively using them. We believe that supposedly unnatural and unreal concepts of digital technology, e.g., tapping or clicking objects, the pinching gesture, undo/redo, or the file system's files and folders have become deeply internalized concepts that are used effortlessly by users when being exposed to new UIs. Although it might be tempting to use the vast design space of post-WIMP interaction (e.g., gestural interaction, TUIs, augmented reality, and organic UIs) to radically redesign the entire face of computing, we should not ignore the already existing building blocks of

the digital world just because they seem to originate from less fortunate uses of real-world metaphors or bad design decisions and technological limitations from the past. *Blended Interaction* is based on blending theory and conceptual integration, which teach us about the benefits of using basic-level sensorimotor concepts during interaction design, but they also explain why we still can learn and apply some complex non-"real-world" concepts from the digital world without an extensive cognitive load.

4.3 Power versus reality tradeoffs

The art of designing good UIs for *Blended Interaction* lies in finding good design tradeoffs. Designers have to find design solutions that are situated in a continuum between two extremes:

1. Expressive UIs that provide direct access to the raw computational power or "magic" of the digital world but rely on hard-to-learn and typically unfamiliar concepts, e.g., command line interfaces, database queries, programming languages, complex editing applications, IRC chats, and raw motion sensor data.
2. Easy-to-learn UIs that make use of familiar real-world concepts but are often inferior in expressive power, precision, versatility, scalability, and efficiency, e.g., multi-touch and TUIs, natural language queries, visual programming languages, simple mobile editing apps, virtual reality chat rooms, and gesture recognizers.

Successful designs of *Blended Interaction* therefore blend the virtues of familiar concepts from the physical or social realm with the expressive power of the digital world in a considered manner, so that ideally only the desired properties of each are preserved while the undesired are discarded. The goal is to preserve the user's perception of familiar and natural interaction with the physical or social realm within an interactive space by creating post-WIMP UIs that carefully blend our familiar reality with the novel powers of computation.

In the context of reality-based interaction, Jacob et al. [24] refer to this design challenge as the *power versus reality* tradeoff. Here, "power" can be seen as a generalization of the great expressive power, functionality, efficiency, and versatility of the digital world. The design goal is to give up "reality" only explicitly and only in return for increasing power. Jacob et al. [24] suggest that adding reality to an UI without loss of power will make them better, and that giving up reality to gain power should be done carefully: "The designer's goal should be to allow the user to perform realistic tasks realistically, to provide additional non-real-world functionality, and to use analogies for these commands whenever possible." Jacob et al. also list typical design tradeoffs to structure discussion and

comparing alternative options during design space analysis, for example *reality versus efficiency* (e.g., interacting with TUIs vs. keyboard hotkeys) or *reality versus versatility* (e.g., single task TUI vs. multi-purpose GUI) [18].

We regard Jacob et al.'s four themes of reality and their considerations about power versus reality tradeoffs as an important basis for *Blended Interaction*. However, as discussed before, *Blended Interaction* does not only explain how the four themes of reality from the physical, non-digital world can be used for better interaction design, but also how other prior experiences including those of digital technology can become a part of users' reality. *Blended Interaction* acknowledges the fact that some concepts from the digital world have been adopted and deeply internalized by the user population and are applied almost as effortlessly as if it were basic-level sensorimotor experiences. Therefore, the notion of "reality" or the "real world" in *Blended Interaction* includes those well-established concepts learned from the experiences of digital technology.

Finding good design tradeoffs is not a new challenge in HCI. For instance, Rosson and Carroll [33: p. 101,186] suggested brainstorming about metaphors and functions during the phases of activity design and interaction design and systematically analyzing their individual pros or cons during *claims analysis*. An example tradeoff they mention is providing the users of an application with familiar templates that already offer typically used components: On the one hand, this simplifies and guides the creative process and builds on prior experience. But on the other hand, it also may discourage more inventive and creative results [33: p. 100]. We agree with Rosson and Carroll that it is important to make such pros and cons explicit, in particular when designing *Blended Interaction*. The "folders" mentioned above can serve us here as an example: Using virtual folders in an interactive space enables users to use familiar hierarchical structures to organize resulting artifacts, for example, the results of a brainstorming or brainsketching session. However, this disallows the use of less familiar but more expressive digital information structures such as user-tagged collections or networks of artifacts that enable to trace design ideas inside the history of many sessions (see for example [34]). Therefore, using folders may result in a more familiar, easy-to-learn UI with a lower threshold, but may also decrease its expressive power and what can be accomplished with it.

A further technique for making the pros and cons of design tradeoffs explicit is PIBA–DIBA ("Physicality is Better At, Digitality is Better at") lists. These lists were invented by Hurtienne et al. [20] for TUI designers and are intended for assigning each function from a system's requirements list to either the digital or physical realm during the design process. For example, in some cases, it is good to represent objects physically, because physical

objects provide instant haptic feedback and are easy to grasp (both mentally and physically). However, in other cases, it is better to choose a digital representation, because digital objects can be created from nothing and as copies of other digital objects [20]. By contrasting the virtues of the physical and digital realm, PIBA–DIBA lists support decision makers during design activities. They help to make the possible design alternatives and their pros and cons visible, and they invite to reflect about hybrid solutions in which a function bridges the physical and digital realm.

We believe that designers of *Blended Interaction* can benefit from these aforementioned approaches, in particular if they are combined with an analysis of UI blends according to the diamond-shaped diagram from Fig. 2. Designers can decompose one or multiple concepts from a UI design (a blend or blends of blends such as the "folder" from Sect. 3) to identify the necessary input spaces (e.g., "physical folder," "file system folder"), generic spaces (e.g., "container," "label"), and projections that have to be present in the users' conceptual systems to arrive at an emergent structure that is similar to the one that the designer had in mind. This helps to review if concepts from a UI design are likely to work for the targeted user population and their assumed prior knowledge. It also can help to relate the different concepts that users are exposed to and assess how well these concepts fit together in an overall coherent structure. These insights can help designers to reflect about and rethink their design and make appropriate shifts in the power versus reality continuum such as integrating functionality in a less familiar but more expressive way or vice versa.

5 Applying Blended Interaction

After introducing the different elements and fundamental principles that our framework is based on, we can now provide a visual overview of *Blended Interaction* (see Fig. 4) that summarizes its defining features and can serve as visual reference during the following sections.

In brief, *Blended Interaction* is based on blends (see Fig. 4, center) between concepts from the users' familiar reality, including already well-established digital concepts (see Fig. 4, left), and the expressive power of digital computation (see Fig. 4, right). Viewing these blends through the lens of conceptual integration with input spaces and generic spaces enables designers to analyze existing designs, generate new design ideas, and to inform design tradeoffs within the four *domains of design* of post-WIMP *Blended Interaction* (see Fig. 5) that we introduce in greater detail below. Understanding conceptual integration and deconstructing UI concepts into blends, input spaces, and generic spaces (see Fig. 2) can help to directly or

Fig. 4 A visual overview of our framework of *Blended Interaction*

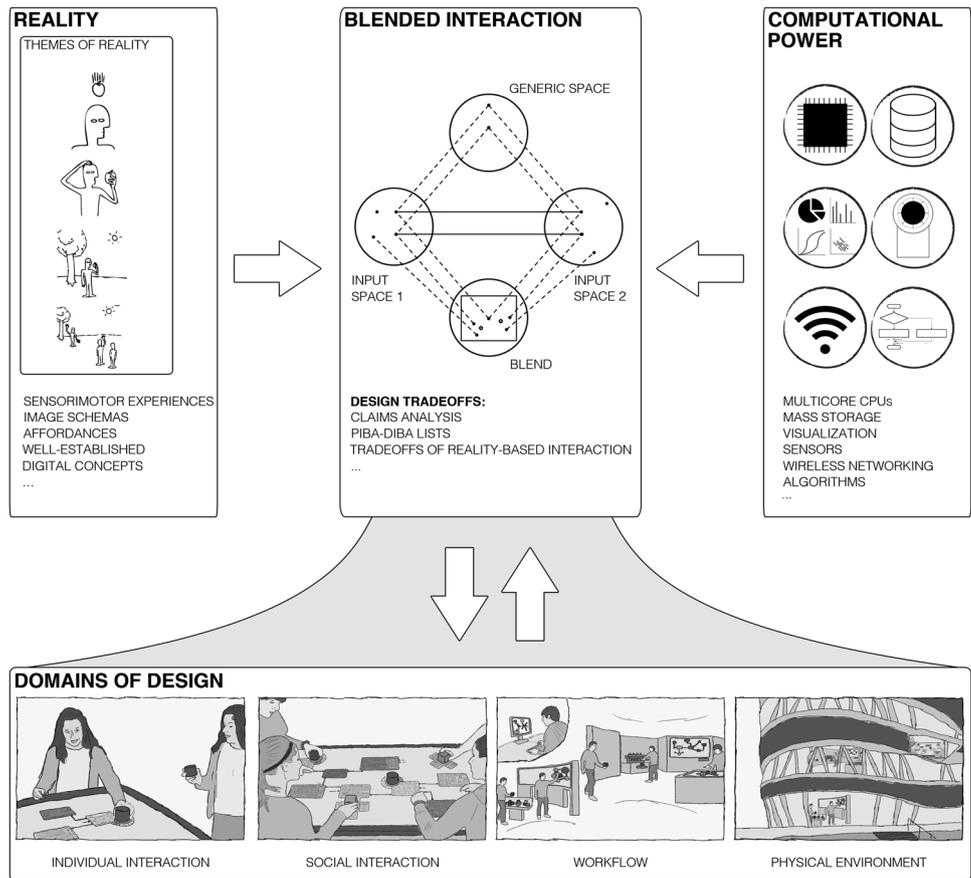
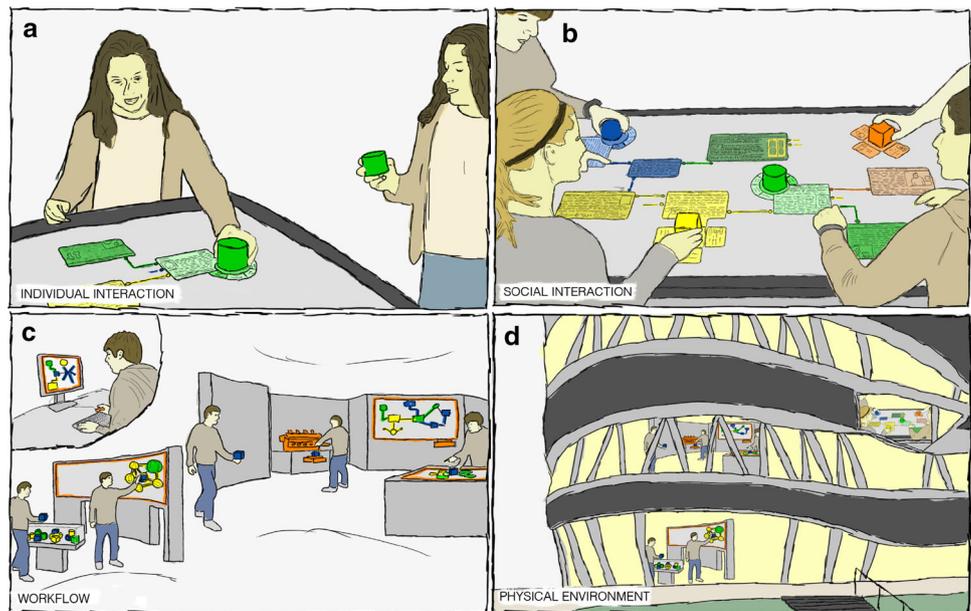


Fig. 5 The four *domains of design* of *Blended Interaction*: **a** individual interaction, **b** social interaction, **c** workflow, and **d** physical environment



indirectly sharpen existing design tools such as claims analysis, PIBA–DIBA lists, or the tradeoffs of reality-based interaction (see Fig. 4, center) and can lead to better design tradeoffs.

In particular, *Blended Interaction* draws strength from considerate blends of what cognitive science and HCI have identified as basic-level experiences and building blocks of human thought (e.g., themes of reality, image schemas, and

affordances) and the “magic” that post-WIMP interfaces and their underlying computational power provide (e.g., high-performance CPUs, high-resolution displays, wireless networks, mass storage, and smart objects). By viewing HCI through the lens of conceptual integration and blends, it becomes obvious that designers should consider such basic-level experiences and building blocks before resorting to elaborate conscious analogies, e.g., elaborate real-world UI metaphors or simulations. Conceptual integration also explains why well-established digital concepts have become a part of users’ reality, and we cannot consider our reality as free of digital influences anymore. Therefore, already well-established digital concepts should be taken into account, even when designing radically new post-WIMP UIs.

5.1 The four domains of design of Blended Interaction

To better structure the problem and design space of post-WIMP interaction, we divide it into four subspaces to which we refer as the four *domains of design of Blended Interaction*. These four domains are based on following considerations:

1. *Individual interaction*: At the heart of every collaborative activity is an individual’s ability to interact with an artifact and to manipulate it. Such individual interactions are the basis of all group activities, and therefore, they must be supported by adequate interaction techniques. In terms of technologies, this can be provided for example by the use of tangible interaction or pen, multi-touch, and multi-modal interfaces instead of traditional mouse and keyboard interaction with navigation and pointing in multiple windows, menus, or pages. For instance, the illustration of individual interaction in Fig. 5 shows how a physical token can be put on an interactive tabletop to display and move information and can also serve as a rotary knob for directly manipulating parameters or content.
2. *Social interaction and communication*: Collaboration needs a coordination of individual activities that can be achieved through social interaction and communication. As is described by Jacob et al. for their theme of *social awareness and skills* [18], users have established social norms, protocols, and practices in their daily lives that they rely on during collaborative activities. When designing interactive spaces, we need to take these into account to avoid interfering with natural social interaction, territoriality, and communication. Furthermore, we must design technologies that support these established norms, protocols, and practices, e.g., by providing technology-supported externalizations of the task at hand such as a visual-tangible representation on a horizontal interactive tabletop that increases group awareness and affords equitable interaction (see Fig. 5).
3. In practice, the structure and sequence of individual and social interactions are often guided by a higher-level *workflow* that defines the typical phases of collaboration and their intended outcomes. Designs that enforce a workflow can be helpful, e.g., when designing a system for a clearly defined creativity technique or business process. However, they can also be disruptive because today’s knowledge-intensive and creative work practices are often situated in an ill-defined social setting without the clear workflows that are typically at the heart of traditional business information systems. In these cases, no UI can provide the structure that is lacking in the task itself [35: p. 35]. It is therefore necessary to find tradeoffs between enforcing clearly defined workflows on users by providing functionality in sequences, templates, or page flows and, on the other hand, creating designs that are open for “unanticipated use” [36] and encourage more inventive and creative results [33: p. 100]. In any case, *Blended Interaction* must be concerned with the efficient support of the users’ actual tasks and work processes and must not be preoccupied with only designing individual and social interactions. This focus on workflow is key to the wider adoption of post-WIMP designs and in our eyes is not sufficiently addressed in post-WIMP literature yet.
4. Last but not least, colocated collaboration happens in a *physical environment* (see Fig. 5), i.e., the physical layout and architecture of the room or work environment, including the room itself, its furniture such as tables, chairs, floors, and ceiling, but also the form factors of digital devices such as display sizes or shapes. This also includes the use of sound and lighting or attempts to control the noise level. In some cases, this physical environment already exists, and only minimal alterations can be done, for example, when creating large blended spaces in outdoor museums and living history villages [37] or in the buildings and streets of our future connected cities. Naturally, understanding and designing the physical environment involves many more disciplines than just HCI, e.g., architecture or industrial and interior design. Although Streit et al. [38] already motivated a more explicit relationship between HCI and the field of architecture in 2001, we feel that HCI typically focuses only on few selected aspects, e.g., size, position, mobility, orientation, readability of interactive displays, and their ergonomic, and social consequences. While this is important to avoid situations in which the physical configuration of digital devices conflicts with natural

collaboration styles, there are many other aspects of the physical environment that are relevant for post-WIMP interactive spaces and that should be addressed in close cooperation with disciplines that research and design the built environment.

5.2 An example of Blended Interaction: AffinityTable

In the following, we outline our framework's contribution to HCI and interaction design and provide examples of how we employed the framework in our work. Our example is an interactive space that we designed for supporting the collaborative design activity of "affinity diagramming" [39]. This design method was originally proposed by Beyer and Holtzblatt [40] as a part of contextual design and serves the exploration and discussion of design ideas or design problems within small groups. The key outcome of the method is a large spatial layout with clusters of design ideas. Traditionally, the method is best practiced using a lo-tech approach involving physical design materials, paper sticky notes, and non-digital whiteboards. Beyer and Holtzblatt [40] even argue against using any digital tools as they may break the workflow and social dynamics that characterize the method. Considering this close relation between affinity diagramming and physical artifacts, designing computational support for the method is a particularly interesting and challenging example for the application of our framework of *Blended Interaction*.

At the beginning of our design process of AffinityTable, we conducted an analysis of traditional, non-digital affinity diagramming to observe and understand natural work practices and to identify design opportunities for blending these practices with computational power. Therefore, we observed the use of material artifacts and physical workspaces and reflected about the physical properties that should be preserved and those that should be augmented or replaced by digital technologies.

For example, we found that individual interaction with paper-based sticky notes is highly effective for idea generation activities and for individual reflection of design ideas. Similarly, we observed how individual and shared territories were used for specific purposes and that any digital technology that interferes with the natural partitioning of space will substantially change the nature of social interaction. We further found that the workflow of affinity diagramming is determined and dependent on a combination of individual and shared physical actions such as grabbing, pointing, skipping, navigating, collecting, and conversing. Eventually, the physical environment in which the method is practiced, such as the room, the available furniture, and display spaces, is inherently defining the social dynamics, productivity, and fluency of the

collaborative design activity. Therefore, we saw great potential for improving aspects such as accessibility, visibility, and fluency of the group work by addressing physical restrictions imposed by spatial and material characteristics with computational power. However, digital technology may not only support the basic workflow of the method, but also rather augment it with new features that are not possible using traditional non-digital tools and materials. The goal of our design phase was hence to design a new, more powerful method that nonetheless builds upon the knowledge and experiences of designers familiar with the original workflow. This resulted in a series of necessary design tradeoffs, which we had to make. In the following, we describe selected examples from the resulting interaction design along the four domains of design of *Blended Interaction* and highlight how the notion of blends guided our design decisions.

5.2.1 Physical environment

In the initial study of traditional affinity diagramming (see Fig. 6, left), we found that the different work surfaces and their specific characteristics and orientations are valuable, and replacing them with typical mobile or desktop devices would disrupt the social dynamics and the workflow. In our analysis, we found that the horizontal table was well suited for individual actions such as writing sticky notes and for collaborative actions such as sorting, skipping, and grabbing them. Similarly, the vertical whiteboard facilitated gaining an overview and increased the visibility of shared notes leading to greater group awareness. Accordingly, our goal was that the basic characteristics of the traditional physical environment in which the activity is usually practiced would be preserved [39]. The design of the physical environment should be informed by our knowledge of the users' *environment awareness and skills* and their *social awareness and skills* [18]. Therefore, the existing non-digital whiteboards and flip charts became a starting point of our design and input spaces for the blends underlying the UI concepts created during our design activity.

At the same time, however, we observed that the need to manually transfer physical artifacts between the workspaces reduced the efficiency of the collaboration. Furthermore, the results of a session can only be preserved by keeping or by recreating the spatial configuration of the artifacts on the whiteboard, which is either very space-consuming or time-consuming. Our solution was to replace the non-digital whiteboard and table with interactive surfaces of similar size and orientation. This resulted in a hybrid workspace with a horizontal digital tabletop. The tabletop is used for displaying, populating, and spatially arranging a workspace containing sticky notes that are

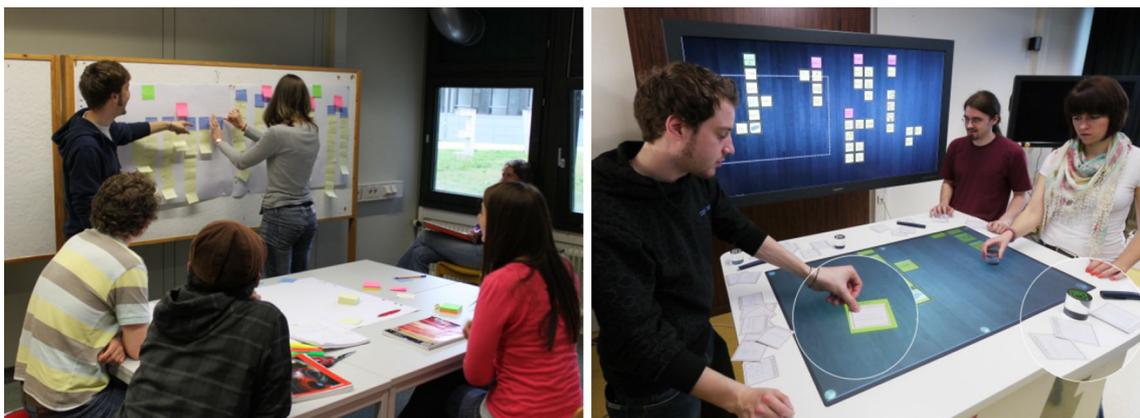


Fig. 6 Non-digital affinity diagramming (*left*). The resulting AffinityTable prototype [39] that consists of a tabletop, a large high-resolution display, digital pen and paper, and tangible user interface elements (*right*)

created using digital pen and paper. Furthermore, a large high-resolution wall display is positioned in close proximity to the table to provide an additional view of the workspace (see Fig. 6, right). Since both surfaces' content is shared and rendered by digital computation, previous states and configurations can be easily preserved and recreated. The resulting surfaces blend the benefits of traditional work surfaces and digital technology and do no longer require a manual transfer of artifacts between both surfaces. This preserves the benefits of digital and non-digital surfaces, but concentrates manipulations on a horizontal surface that is equally accessible to all participants. Because the vertical display cannot be blocked, it further improves the visibility of artifacts.

5.2.2 Social interaction and communication

In the case of the AffinityTable, our design goal within this domain was primarily to augment coordination and increase group awareness by providing adequate interaction techniques and visualizations. For example, our analysis of non-digital practices revealed that the size of the sticky notes leads to a lack of visibility and mutual blocking during collaborative activities. Although our digital overview display already contributes to the goal of increasing the visibility of shared artifacts, they still remain rather small when viewed from a distance, and hence, group awareness is still hampered (see Fig. 6, right). To address this issue, AffinityTable introduces a physical “focus token” (see Fig. 7). By placing and moving this token on the table, a rectangular region within the virtual workspace is selected for a detailed view on the vertical display (see Fig. 7, top). By rotating the token (see Fig. 7a–c), users can increase or decrease the zoom factor of the magnified view on the vertical display. By lifting the token from the table, the original view is restored. Because

all other artifacts are temporarily hidden when using the token (except the immediate surroundings on the table), this technique is a powerful tool that allows placing focus on selected artifacts in order to facilitate awareness during discussion and reflection. It also minimizes physical movements and mutual blocking. Accordingly, it uses features of digital functionality to address limitations of the real world in a fluent and accessible way.

The focus token blends different inputs from the non-digital realm to enable users to effortlessly control the content of the vertical display and coordinating collaboration:

First, the token is important as an indicator, since the user who is currently holding or using the token can be clearly identified as the person who is currently in charge. Holding the token shows who is currently in control and makes other group members aware of this. This facilitates turn taking and coordinating collaboration. Users are familiar with this principle of passing around an object to coordinate collaborative activities. For example, we know this from passing around a microphone in the audience during a Q&A session or handing over a dice when playing a board game. Kindergarten children are taught to pass around a “talking stone” for turn taking during talking circles, a concept based on the “talking stick” that is used in tribal council circles by Native American tribes. Although the focus token greatly differs from these examples in its visual appearance, the underlying principle is familiar and unambiguous to understand.

Second, the presence of the token on the table visually indicates that the system is now in a different mode until the token is removed again. This increases the group's awareness for the current system mode and also clarifies how switching modes can be done, namely by simply removing the token from the tabletop or putting it back.

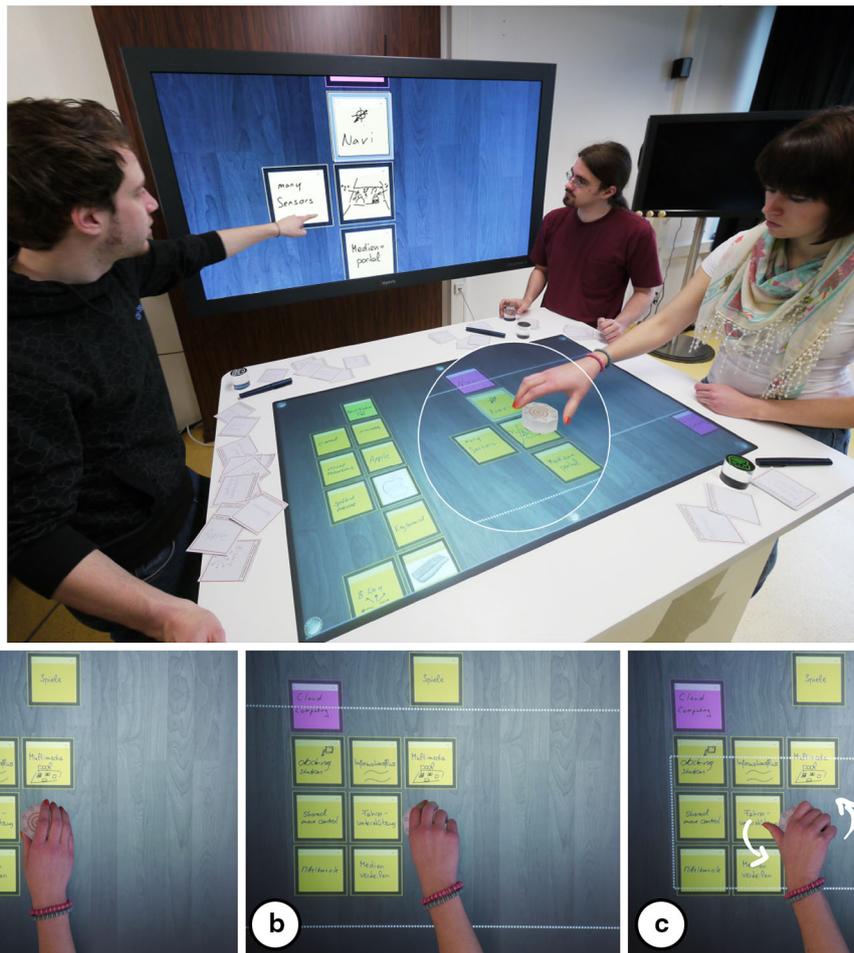


Fig. 7 The physical “focus token” (top) is a tool for zooming into a specific area of the virtual workspace. By moving and rotating the token (bottom, a–c), users can move and resize a rectangular section

on the tabletop that is then displayed in great detail on the vertical high-resolution display [39]

Third, the token serves as a controller and resembles a rotary knob known, for example, from hi-fi or car audio equipment. Its underlying principle has become a part of our everyday experience: Clockwise rotation increases an effect (e.g., greater volume, size, and magnification), while moving counter-clockwise reduces an effect (e.g., less volume, size, and magnification). Users can easily understand and apply this way of controlling the size of region that is displayed on the vertical display.

5.2.3 Workflow

During our design of AffinityTable, we closely studied existing work practices [39]. We identified three phases of affinity diagramming: (1) Users work on their own in individual workspaces to create ideas, (2) users present their ideas and share them, and (3) users collaboratively reflect and cluster ideas. For each phase, we created interaction techniques and UI concepts that supported the users’ tasks

based on blending reality with digital expressive power. Thereby, we also referred to themes of reality such as *naïve physics*, *body awareness and skills*, and *environment awareness and skills* [18] as input spaces to design interaction techniques that support the basic workflow of the design method but extend it with novel functionality that can only be realized using computational power.

One example of these UI concepts is the “hybrid sticky note” that blends properties of physical sticky notes with that of digital objects. Notes can be generated with digital pen & paper technology (see Fig. 8, left) and copied to the shared workspace by placing them on the interactive table (see Fig. 8, right). A digital representation of the physical note appears within the virtual workspace (on both displays), while the physical copy can be kept in the personal workspace. They hence support easy generation and keeping of notes (phase 1) by writing on physical paper (see Jacob et al.’s *body awareness and skills*). Furthermore, it uses the “magic” of the digital world to let users spatially



Fig. 8 The hybrid sticky note: Users create sticky notes using digital pen and paper. They can keep the physical originals (*left*) and also paste a copy of them into the shared workspace by putting them on the tabletop (*right*)

arrange and share a digital representation of the content in the workspace and on multiple displays (phase 2). This is related to Jacob et al.'s theme of *environment awareness and skills*.

Users may manipulate their notes later on by writing or crossing out words on the physical note. These manipulations are synchronized with the digital copy. When placing a note onto the table that was already copied into the workspace, it is not duplicated but moved to this new location instead. Overall, hybrid sticky notes allow the sharing of artifacts with minimized transfer costs while still preserving the ability to ideate using physical material and the ability to reflect upon all individual contents in the personal workspace throughout the session. It also enables rapid switching between the personal and the shared workspace. Once transferred into the shared workspace, the digital note has characteristics that go well beyond what is possible with traditional material. For example, interaction techniques for more efficient clustering, piling, collecting, and gesturing were designed for supporting the phase of reflecting and clustering (phase 3).

AffinityTable uses different blends to provide efficient techniques for clustering digital representations that make use of our familiarity with physical behaviors (see Jacob et al.'s [18] *naïve physics*). For example, a simple simulation of physical attributes is provided with the ability to drag, rotate, and flick digital notes by using typical multi-touch manipulations on the interactive table. Ideally, users quickly learn to handle the digital objects with the same fluency and expertise that they have when they handle traditional non-digital materials.

Furthermore, to facilitate the collaborative organization of notes into groups, we included a simple clustering algorithm that automatically aligns and associates notes when released close to each other (see Fig. 9a–c). In the physical realm, such a magnetism-like behavior does not exist for physical notes and they have to be aligned manually.

Nevertheless, the concept of objects exerting force on each other, for example through gravitation or magnetism, and thus moving closer and eventually snapping and sticking together is familiar to users and also part of *naïve physics*. Similarly, piles of notes can be created by releasing notes on top of other notes, so that they are piled but are aligned automatically. Entire clusters and piles can be moved by dragging them with multiple fingers (see Fig. 9d–f).

Eventually, for supporting collaborative reflection (phase 3), a further interaction technique was introduced that supports the retrieval of images based on the content of the notes. It enables users to quickly retrieve digital images from an internet image search service using the handwritten text in the notes as textual search query. This can be useful to search for inspiration during phases of reflection or ideation. Once notes with the text to search are placed into a search region on the table, the found images appear underneath the note's position. The note, but also the images, can then be added to the shared clusters to augment them with representative images. This concept is an example of *Blended Interaction* that draws from already well-established digital concepts. We believe that in this case, the key elements of the input space “search engine” such as “search term” and “results” are very well established among the user population so that any alternative analogies or designs would rather lead to confusion than to a more natural interaction.

5.2.4 Individual interaction

In the case of the AffinityTable, a good example of blended *individual interaction* is the hybrid sticky note based on blending physical paper and pens with the virtues of a digital representation. By providing users with their own notes and pens, users are able to effectively ideate and reflect in personal workspaces individually without having to interfere with other group members, for example,

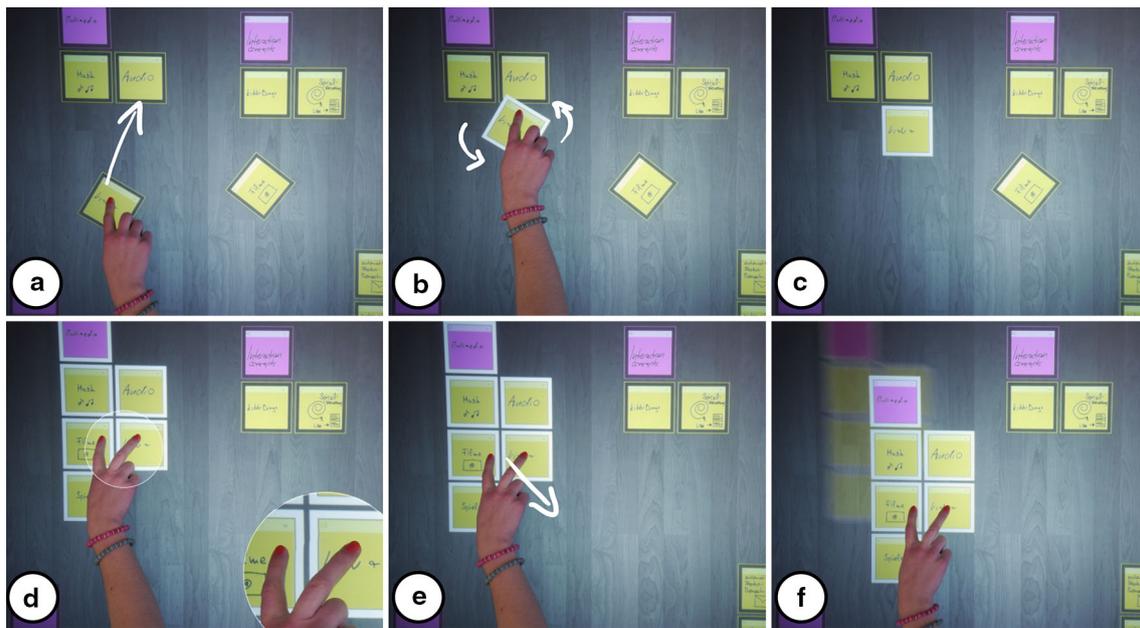


Fig. 9 Clustering digital notes. Associating notes with clusters (a–c) and moving clusters with multiple fingers (d–f) [39]

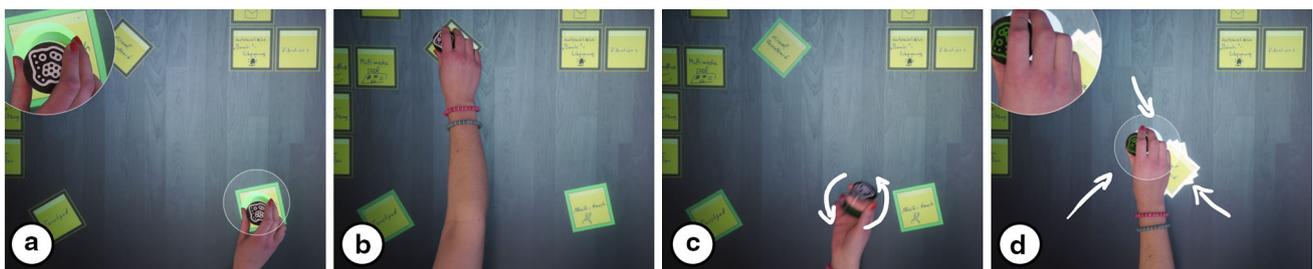


Fig. 10 The “collector token” enables collecting and moving multiple notes within the digital workspace [39]

because of having to share input devices such as a keyboard or a digitizer tablet.

Another example is the “collector token” that we designed to support users who have to interact with multiple notes. It is based on the observation that users frequently have to select notes from different parts of the workspace and to move them to a new location for creating a new cluster or pile. Manually dragging multiple notes to a new location can take a long time and can become tedious. Therefore, the collector token, which is personalized and color-coded for each user, lets a user select multiple individual notes from the virtual workspace. Similar to a rubber stamp, the token is put on each note to select. After “stamping” a note, it becomes highlighted and is colored with the color of the token (see Fig. 10a, b). Furthermore, the token also behaves like a container in which all selected notes are collected. By turning around the token and putting it back on the table’s surface, the container is emptied and the selected notes are poured out of the container to the new location (see Fig. 10c, d).

5.3 Contributions and limitations of our framework

As the above examples demonstrate, an understanding of the physical environment, the workflows, social interaction, and communication as well as individual interaction can contribute to a holistic system design that truly supports the complex nature of collaboration and preserves natural work practices. Thereby viewing these domains of design through the lens of *Blended Interaction* can help designers to preserve the virtues of familiar reality and the expressive power of the digital world and to create more usable and consistent systems. The success of our AffinityTable in user studies reveals the practical value that *Blended Interaction* had as a point of orientation for us during our design. However, we do not claim that others need *Blended Interaction* to achieve similar results. We also do not claim that our framework already provides a design method or a step-by-step recipe for interaction design that guarantees a high-quality outcome. It is important to understand that *Blended Interaction* in its

current stage is an *explanatory* and not a *prescriptive* conceptual framework.

According to Rogers, who provides a comprehensive account of classical, modern, and contemporary theory in HCI in [41], the most prolific and, arguably, successful developments in HCI theory have been conceptual frameworks that, for example, are derived from an imported theory (in our case blending theory). “Conceptual frameworks can vary along a continuum of prescription explanation: the more prescriptive a framework the more likely it will consist of a series of steps or principles to be followed. The more explanatory a framework, the more likely it will consist of a set of concepts or dimensions to be considered” [41: p. 82]. Rogers also writes that imported theories have been least successful when adapted as generalizable methods (i.e., the prescriptive and predictive categories), intended to be used by practitioners. She considers it as foolish to assume or hope that theories “do design.” Their input can only ever be indirect, e.g., by providing concepts and analytic tools. “A theory cannot provide prescriptive guidance in the sense of literally telling a designer what and how to design” [41: p. 84].

Here, our intention behind *Blended Interaction* is to provide such indirect input with an explanatory framework that does not try to tell designers *how to* design but to explain them *why* some designs are more successful than others. Therefore, *Blended Interaction* stands in the tradition of similar work such as the cognitive account of *direct manipulation* by Hutchins et al. [9] or Jacob et al.’s *reality-based interaction* [18]. We hope that after introducing our framework to the HCI community, researchers and practitioners will start to pick up our explanations and report about their benefit during their work. As Rogers discusses, evidence of the success of conceptual frameworks can be counted in the reporting by others of having used them in different projects, ideally case studies of design practice [41: p. 82]. Given the fact that *Blended Interaction* has already been adopted by user experience design studios that specialize in applied research,⁴ we hope that the future will provide us with such reports to support our claim. These reports about the experiences of designers will also enable us to develop *Blended Interaction* further and to extend it beyond a purely explanatory framework to a more prescriptive framework that also provides direct recommendations in the sense of a design methodology.

⁴ Blended Interactions user experience design studio at Rochester Institute of Technology. <http://blendedinteractions.com/about/> (Last accessed Jul 19, 2013).

6 Conclusion

In this article, we introduced *Blended Interaction*, a new conceptual framework that helps to explain when users perceive UIs in post-WIMP interactive spaces as “natural” or not. We believe such a framework is necessary since HCI still lacks a deeper understanding of why some novel interfaces to ubiquitous computing can be used without “mental gymnastics” and are successfully becoming “invisible” [3] and others fail. Therefore, our theoretical framework of *Blended Interaction* is based on a novel and more accurate description of the nature of HCI that consolidates and extends previous and related work from HCI [12, 15, 18–20, 24], cognitive science [13], and cognitive linguistics [14, 16, 17].

At the heart of *Blended Interaction* is an embodied view of cognition with its notion of conceptual integration. Conceptual integration is a mental capacity that we use in almost every moment of thinking without being consciously aware of it [14]. Conceptual integration enables us to create and apply new concepts and explains why we can create and use UIs in HCI that blend experiences of our familiar reality with the unfamiliar expressive power and “magic” of the digital world. It also explains why, generally speaking, concepts of a computer UI that are closer to familiar, everyday bodily, spatial, or social experiences are easier to learn and to use.

In comparison with existing frameworks or core interface techniques, *Blended Interaction* extends our understanding of a “natural” HCI in post-WIMP interactive spaces by four key contributions:

First, in the light of conceptual integration, interaction designers should consider using and blending the vast amount of concepts (e.g., image schemas, themes of reality, and affordances) that we as humans share due to the similarities of our bodies, our early upbringing, and our sensorimotor experiences of the world before resorting to elaborate conscious analogies such as the desktop metaphor. Domain-specific or culture-specific analogies such as “THE OPERATING SYSTEM IS AN OFFICE DESKTOP” that are based on the recognition of real-world objects such as “folders” or “windows” are often not necessary or even disrupting. Similar to *embodied interaction* [15], instead of drawing on artifacts in the everyday world, *Blended Interaction* draws on the way the everyday world works and the ways we experience the everyday world.

Second, however desirable such a clean separation may be, we cannot consider our “reality” as free from digital influences anymore. Even though embodied cognition emphasizes how strongly our bodies and brains shape our cognition and how they are a result of millions of years of evolution, it also considers our cognitive abilities as socio-

culturally determined by our prior experiences. This inevitably includes our experiences with the computers and digital technologies that surround us. Therefore, we should not ignore the already existing building blocks of the digital world just because they seem to originate from less fortunate uses of real-world metaphors or bad design decisions and technological limitations from the past.

Third, the art of designing good UIs for *Blended Interaction* lies in finding good design tradeoffs. Designers have to find design solutions that are situated in a continuum between two extremes: Expressive UIs that provide direct access to the raw computational power or “magic” of the digital world but rely on hard-to-learn and typically unfamiliar concepts and easy-to-learn UIs that make use of familiar real-world concepts but are often inferior in expressive power, precision, versatility, scalability, and efficiency. Deconstructing UI concepts into blends, input spaces, and generic spaces helps to sharpen existing design tools such as claims analysis [33], PIBA–DIBA lists [20], or the tradeoffs of reality-based interaction [18, 24] and can lead to better design tradeoffs by making the pros and cons of a design more explicit.

Fourth, *Blended Interaction* introduces four domains of designs that designers should keep in mind while designing post-WIMP interactive spaces: (1) The design of the *individual interaction* that is the basis of all group activities and therefore must be supported by adequate interaction techniques, for example, by tangible interaction or multi-touch or multi-modal interfaces instead of traditional mouse and keyboard interaction. (2) The design of *social interaction and communication*, so that the interaction design does not interfere with established social norms, protocols, and practices that users rely on during collaborative activities and their coordination. (3) The design of *workflows* to ensure that post-WIMP interaction efficiently supports the users’ actual tasks and work processes without discouraging more inventive and creative results. (4) The design of the *physical environment*, i.e., the architecture of the room or work environment, including the room itself, its furniture such as tables, chairs, floors, and ceiling, but also the form factor of digital devices such as display size or shape.

Acknowledgments The authors would like to thank Johannes Zagermann and Daniel Klinkhammer for providing some of the illustrations and figures for this article. Furthermore, the authors would like to thank all the organizers and participants of our DCIS 2012 workshop “Designing Collaborative Interactive Spaces” at AVI 2012⁵ and our workshop “Blended Interaction: Envisioning Future Collaborative Interactive Spaces” at CHI 2013⁶ for the valuable feedback on and productive discussion of Blended Interaction that has helped to further develop our framework.

⁵ <http://hci.uni-konstanz.de/dcis/>.

⁶ <http://hci.uni-konstanz.de/blendedinteraction2013/>.

References

- Ju W, Leifer L (2008) The design of implicit interactions: making interactive systems less obnoxious. *Des Issues* 24(3):72–84. doi:10.1162/desi.2008.24.3.72
- Oulasvirta A (2008) When users “do” the Ubicomp. *Interactions* 15:6. doi:10.1145/1340961.1340963
- Weiser M (1991) The computer for the 21st century. *Sci Am* 3:94–104. doi:10.1145/329124.329126
- Whittaker S, Terveen L, Nardi BA (2000) A reference task agenda for HCI. In: Carroll JM (ed) *Human–computer interaction in the new Millennium*. Addison-Wesley, New York, pp 167–190
- Byrne MD (2003) Cognitive architecture. In: Julie AJ, Andrew S (eds) *The human–computer interaction handbook*. L. Erlbaum Associates Inc., Hillsdale, NJ, pp 97–117
- Kieras D (2003) Model-based evaluation. In: Julie AJ, Andrew S (eds) *The human–computer interaction handbook*. L. Erlbaum Associates Inc., Hillsdale, NJ, pp 1139–1151
- Blackwell AF (1998) *Metaphor in Diagrams*. PhD Thesis, University of Cambridge, Cambridge
- Jetter H-C, Leifer S, Gerken J, Schubert S, Reiterer H (2012) Does (multi-)touch aid users’ spatial memory and navigation in ‘panning’ and in ‘zooming and panning’ UIs? In: *Proceedings of International Working Conference on Advanced Visual Interfaces (AVI’12)*. ACM, New York, NY, pp 83–90
- Hutchins EL, Hollan JD, Norman DA (1985) Direct manipulation interfaces. *Hum Comput Interact* 1:311–338
- Ravasio P, Tschertner V (2007) User’s theories of the desktop metaphor or why we should seek metaphor-free interfaces. In: Kaptelinin V, Czerwinski M (eds) *Beyond the desktop metaphor: designing integrated digital work environments*. MIT Press, Cambridge, MA, pp 265–294
- Constantine L, Lockwood L (1999) *Software for use: a practical guide to the models and methods of usage-centered design*. Addison Wesley, Reading, MA
- Imaz M, Benyon D (2007) *Designing with Blends: conceptual Foundations of Human-Computer Interaction and Software Engineering*. MIT Press. doi:citeulike-article-id:4207163
- Gibbs RW (2006) *Embodiment and cognitive science*. Cambridge University Press, Cambridge. doi:10.2277/0521811740
- Fauconnier G, Turner M (2002) *The way we think: conceptual blending and the mind’s hidden complexities*. EUA Basic Books, New York
- Dourish P (2004) *Where the action is: the foundations of embodied interaction*. First MIT Press Paperback Edition edn. MIT Press
- Lakoff G, Johnson M (1999) *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. Basic Books, New York
- Lakoff G, Johnson M (1980) *Metaphors we live by*. University of Chicago Press, Chicago
- Jacob RJK, Girouard A, Hirshfield LM, Horn MS, Shaer O, Solovey ET, Zigelbaum J (2008) Reality-based interaction: a framework for post-WIMP interfaces. In: *Proceedings of the SIGCHI conference on human factors in computing systems (CHI ‘08)*. ACM, New York, NY, pp 201–210
- Hurtienne J, Israel JH (2007) Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In: *Proceedings of the 1st international conference on tangible and embedded interaction (TEI ‘07)*. ACM, New York, NY, pp 127–134
- Hurtienne J, Israel JH, Weber K (2008) Cooking up real world business applications combining physicality, digitality, and image schemas. In: *Proceedings of the 2nd international conference on Tangible and embedded interaction (TEI ‘08)*. ACM, New York, NY pp 239–246

21. Blackwell AF (2006) The reification of metaphor as a design tool. *ACM Trans Comput–Hum Interact* 13(4):490–530. doi:[10.1145/1188816.1188820](https://doi.org/10.1145/1188816.1188820)
22. Ba Nardi, Zamer CL (1993) Beyond models and metaphors: visual formalisms in user interface design. *J Vis Lang Comput* 4:5–33. doi:[10.1109/HICSS.1991.184010](https://doi.org/10.1109/HICSS.1991.184010)
23. Bederson BB, Hollan JD, Perlin K, Meyer J, Bacon D, Furnas GW (1996) Pad++: a zoomable graphical sketchpad for exploring alternate interface physics. *J Vis Lang Comput* 7:3–32
24. Jacob RJK, Girouard A, Hirshfield LM, Horn MS, Shaer O, Solovey ET, Zigelbaum J (2007) Reality-based interaction: unifying the new generation of interaction styles. In: CHI '07 extended abstracts on human factors in computing systems (CHI EA '07). ACM, New York, NY pp 2465–2470
25. Boden MA (1994) Précis of the creative mind: myths and mechanisms. *Behav Brain Sci* 17(03):519–531. doi:[10.1017/S0140525X0003569X](https://doi.org/10.1017/S0140525X0003569X)
26. Constantine L (1998) Use and misuse of metaphor. <http://www.foruse.com/articles/metaphor.pdf>. Accessed Feb 5, 2013
27. Norman DA (2002) *The design of everyday things*. Basic Books, New York
28. Andrews C, Endert A, North C (2010) Space to think: large high-resolution displays for sensemaking. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI '10). ACM, New York, NY, pp 55–64
29. Kirsh D (1995) The intelligent use of space. *Artif Intell* 73(1–2):31–68. doi:[10.1016/0004-3702\(94\)00017-u](https://doi.org/10.1016/0004-3702(94)00017-u)
30. Scott SD, Carpendale S, Inkpen KM (2004) Territoriality in collaborative tabletop workspaces. In: Proceedings of the 2004 ACM conference on computer-supported cooperative work (CSCW '04). ACM, New York, NY, pp 294–303
31. Jetter H-C, Gerken J, Zöllner M, Reiterer H, Milic-Frayling N (2011) Materializing the query with facet-streams: a hybrid surface for collaborative search on tabletops. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI '11). ACM, New York, NY, pp 3013–3022
32. Gerken J, Jetter H-C, Schmidt T (2010) Can “touch” get annoying? In ACM international conference on interactive tabletops and surfaces (ITS '10). ACM, New York, NY, pp 257–258
33. Rosson MB, Carroll JM (2002) *Usability engineering: scenario-based development of human–computer interaction*. Morgan Kaufmann series in interactive technologies, 1st edn. Academic Press, San Francisco
34. Geyer F, Budzinski J, Reiterer H (2012) IdeaVis: a hybrid workspace and interactive visualization for paper-based collaborative sketching sessions. In: Proceedings of the 7th Nordic conference on human–computer interaction: Making Sense Through Design (NordCHI '12). ACM, New York, NY, pp 331–340
35. Collins D (1995) *Designing object-oriented user interfaces*. Benjamin Cummings, Redwood City, CA
36. Robinson M (1993) Design for unanticipated use. In: Proceedings of the third conference on European conference on computer-supported cooperative work (ECSCW'93). Kluwer Academic Publishers, Norwell, MA, pp 187–202
37. O'Keefe B, Slutsky B, Iulucci N, Nalbandian A, Thanedar A, Mokey S, Mival O (2013) Mobile experiences for tourism: brick city tours. In: CHI '13 Extended abstracts on human factors in computing systems (CHI EA '13). ACM, New York, NY, pp 1413–1418
38. Streitz N, Tandler P, Müller-Tomfelde C, Konomi S (2001) Roomware: towards the next generation of human–computer interaction based on an integrated design of real and virtual worlds. In: Carroll JM (ed) *Human–Computer Interaction in the New Millennium*. Addison Wesley, Reading, MA, pp 551–576. doi:[citeulike-article-id:10563544](https://doi.org/citeulike-article-id:10563544)
39. Geyer F, Pfeil U, Höchtl A, Budzinski J, Reiterer H (2011) Designing reality-based interfaces for creative group work. In: Proceedings of the 8th ACM conference on creativity and cognition (C&C '11). ACM, New York, NY, pp 165–174
40. Beyer H, Holtzblatt K (1998) *Contextual design: defining customer-centered systems*. Morgan Kaufman, San Francisco
41. Rogers Y (2012) HCI theory: classical, modern, and contemporary, vol Lecture #14. *Synthesis Lectures on Human-Centered Informatics*. Morgan and Claypool