Shared Mental Model for Natural Human-Computer Interaction

Authors: Zhang Liang, Wang Xiaoyu, Liu Fang, Ma Cuixia, Hong'an Wang, Cuixia Ma

Date: 2024-08-18T00:00:00+00:00

Abstract

Classic cognitive models such as the Human Processor have provided theoretical foundations for human-computer interaction over the past four decades. As technology has advanced, human-computer interaction is no longer confined to graphical user interfaces, and innovative natural human-computer interaction modalities have continued to emerge. As a trend for future development, natural human-computer interaction urgently necessitates new theoretical models to guide its design and evaluation. This paper proposes a Shared Mental Model, building upon existing cognitive models and grounded in the characteristics of natural human-computer interaction. This model enhances the three classic modules of perception, cognition, and action, incorporates important findings from social cognition such as Theory of Mind, and introduces a shared space module. The Shared Mental Model can be employed to describe and evaluate natural human-computer interaction processes, and endeavors to provide new theoretical guidance and development recommendations for future interaction modalities.

Full Text

Preamble

A Shared Mental Model for Natural Human-Computer Interaction

Zhang Liang[1],[2], Wang Xiaoyu[1],[2], Liu Fang[3],[4], Ma Cuixia[3],[4], Wang Hongan[3],[4]

[1](Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)
[2](Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)
[3](Key Laboratory of System Software and State Key Laboratory of Computer Science, Beijing Key Laboratory of Human-Computer Interaction, Institute of

Software, Chinese Academy of Sciences, Beijing 100190, China)
[4](School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract

Classical cognitive models such as the Model Human Processor have provided the theoretical foundation for human-computer interaction for nearly four decades. With technological advances, HCI is no longer limited to graphical user interfaces, and creative natural interaction methods continue to emerge. As a future trend, natural HCI urgently requires new theoretical models to guide its design and evaluation. This paper proposes a Shared Mental Model based on existing cognitive models and the characteristics of natural HCI. The model refines the classical three-module structure of perception, cognition, and action, incorporates important findings from social cognition such as theory of mind, and introduces a shared space module. The Shared Mental Model can describe and evaluate natural HCI processes and attempts to provide new theoretical guidance and development recommendations for future interaction paradigms.

Keywords: Natural Human-Computer Interaction, Cognitive Model, Perception, Cognition, Action

Funding: Major Program of the National Natural Science Foundation of China (T2192932); National Natural Science Foundation of China (62272447); Beijing Natural Science Foundation (4212029); National Key R&D Program of China (2016YFB1001201). Zhang Liang and Wang Xiaoyu contributed equally to this work. Corresponding author: Ma Cuixia, E-mail: cuixia@iscas.ac.cn.

Human-computer interaction represents the bidirectional exchange of information between humans and computers, as well as a discipline dedicated to the design, evaluation, and implementation of interactive computing systems and related phenomena. In the context of human-machine hybrid intelligence, humans and machines must engage in increasingly frequent interaction and communication. HCI technology is not only critical for optimizing human-machine hybrid intelligence but also constitutes a core competitive advantage in terminal and application innovation in the intelligent era. We envision that such frequent interactions should be natural interactions. The term “natural” in Natural Human-Computer Interaction refers to interactions that are “spontaneous,” “intuitive,” and emerge naturally without excessive cognitive effort [1]. While novel interaction methods continue to emerge in the natural interaction domain, both design standards and theoretical frameworks remain nascent. Currently, there is no single, consistent definition of “natural HCI” [2], though experts in the field share similar perspectives. Unlike the precise, discrete user input characteristic of graphical user interfaces, natural user interfaces provide more natural interaction methods that enable users to interact with computers as they would with real-world objects [3, 4]. Researchers consistently emphasize the importance of human experience in natural HCI [4, 5]. When "naturalness" defines product or interface characteristics, it does not refer to a specific interface type or interaction technology but rather to human experience—the psychological feelings users have when using products and interacting with interfaces. Natural user interfaces aim to reduce cognitive load, shorten learning curves, and optimize user experience based on human psychological feelings and behavioral characteristics [6]. Achieving this requires understanding the psychological processes, characteristics, and patterns that occur when users interact with natural interfaces. Traditional HCI research has proposed cognitive models to describe or simulate human behavior to guide design and evaluation. However, these classical cognitive models were developed for conventional HCI, and natural HCI possesses unique characteristics that demand new theoretical guidance.

1.1 The Model Human Processor and Its Evolution

Traditional cognitive models can be used to evaluate interaction operation times, analyze interaction processes, and substitute for users in simulations. In 1983, Card proposed the Model Human Processor (MHP) to concisely and efficiently describe and predict HCI processes [7]. The MHP model summarizes human psychological processing into three processors: perception, cognition, and action. Building upon MHP, researchers have developed more refined models to simulate human behavior, such as the GOMS model [7] and ACT-R model [8]. The GOMS model simulates the knowledge and cognitive processes users employ when interacting with systems, consisting of four components: Goals (what the user wants to achieve), Operators (basic actions the user performs), Methods (sequences of operators to accomplish goals or sub-goals), and Selection Rules (rules for choosing among alternative methods). GOMS is the most widely used information processing model in HCI and interface design, particularly in usability testing. HCI tasks often involve multitasking. To address behavioral simulation in multitasking scenarios, researchers combined queueing network models with MHP to propose the QN-MHP model [9].

With the development of artificial intelligence and HCI, new demands have emerged for cognitive models. Researchers have proposed an extended GOMS model based on traditional GOMS [10], comprising Goals (user objectives and understanding of those objectives), Operators (actions users perform to achieve goals), Methods (multimodal information during user interaction), and Synthetization (integration of multimodal information in specific contexts). The extended GOMS proposes an evaluation function f(A, C, E) where A refers to Attention (requiring minimal attentional resources), C refers to Cognitive Load (the lower the better), and E refers to Effectiveness of Interaction (including naturalness, directness, and operation speed). Psychologists and computer scientists have proposed a psychological model of human-computer cooperation for the intelligent era to better guide HCI design [11]. This model suggests that human-computer interaction is essentially human-human interaction, possessing similar attributes and patterns.

1.2 Applications and Limitations of Traditional Cognitive Models in HCI

A primary application of cognitive models in HCI is simulating human operation behavior and predicting interaction times. Card and colleagues established time parameters for the three processors of perception, cognition, and action, enabling rough estimates of task execution times [8]. The QN-MHP computational model can simulate human multitasking operations; for example, it can effectively predict driver behavior models and operation times during vehicle driving and map reading tasks [12].

Traditional cognitive models focus on simulating and predicting human behavior to inform HCI design. The psychological model of human-computer cooperation proposes that HCI is essentially human-human interaction, attempting to address more complex interaction processes in the intelligent era. However, previous models have modeled the three main components of perception, cognition, and action from an individual perspective, neglecting individual differences and developmental changes in HCI over time. HCI technology continues to evolve, and the integration between humans and machines is increasing. Particularly in human-machine hybrid intelligence, humans must understand how machines perceive the world, make decisions, and cooperate seamlessly with them [13]. Therefore, the human-computer interaction process requires not only analyzing information perception, processing, and action from an individual perspective but also considering the two as an integrated whole [14].

Furthermore, the HCI field has evolved beyond traditional desktop metaphors, and interaction environments have become increasingly complex and variable. Different environments affect the HCI process and influence individual information perception and action execution. For instance, in noisy environments, both humans and computers must increase the volume of auditory signals. Cognitive models for the natural HCI stage must consider not only individual perception, cognition, and action but also the multiple processes involved in human-machine interaction.

2 A Shared Mental Model for Natural Human-Computer Interaction

Based on cognitive models such as MHP and the psychological model of human-computer cooperation for the intelligent era, and incorporating the goals of natural HCI, we propose the Shared Mental Model, illustrated in Figure 1 [Figure 1: see original paper].

Figure 1. Schematic diagram of the Shared Mental Model

The model operates on five fundamental assumptions:

Assumption 1: The most natural HCI can be equated to human-human interaction, where both humans and machines are individuals connected to other components of the interactive system.

Assumption 2: Users and other components of the interactive system each possess individual processing spaces for information input, processing, and output, comprising basic perception, cognition, and action modules.

Assumption 3: Users and other components of the interactive system exist in both a shared external space and a shared internal space. The shared external space refers to the common context where individuals perceive information, process cognitions, and select and execute actions. The shared internal space refers to overlapping portions of individual interaction spaces formed through evolution or prior experience, encompassing shared experience modules and theory of mind modules.

Assumption 4: Individual processing spaces and shared spaces are continuously updated. The shared space can expand the shared experience system and strengthen the theory of mind module through the HCI process. Simultaneously, optimization of the shared space enables optimization of internal processes within individual processing spaces.

Assumption 5: The naturalness of HCI is primarily determined by the shared space, specifically by the size of the internal shared space (the extent of the shared experience module and the sophistication of the theory of mind) and the similarity of information understanding perceived by different individuals in the shared external space, while also being influenced by the processing efficiency of individual processing spaces.

2.1 Individual Processing Space

Each user or complete computer system can be considered an individual. Information reception, processing, and action execution for each individual occur within their respective individual processing space, which is divided into multimodal perception, cognitive, and action modules.

(1) Multimodal Perception Module

The multimodal perception module transforms external raw information into internal representations that the cognitive module can process. It acquires external information through multiple single channels and integrates information across channels via a multimodal perception module to derive consistent, complete representations.

Individuals in interactive systems possess multiple sensory channels for information acquisition. Humans have five basic senses—vision, hearing, smell, touch, and taste—plus other perceptual modalities such as proprioception, thermoception, and time perception [15]. Computers can acquire images, speech, sensor data, EEG signals, and other multichannel information. Computer perception channels require continuous enrichment to match the diverse sensory channels of humans.

The human brain continuously receives information from various channels and integrates it [16]. Different sensory channels influence one another. For example, cognitive psychology experiments have demonstrated that visual channels can affect auditory information processing (the McGurk effect [17]). Computers should similarly process information from all channels jointly, considering inter-channel integration. Incorporating multisensory integration in interaction will create novel products and substantially enhance interaction experience [18]. However, information across channels must be consistent; otherwise, it may cause misunderstanding in receiving individuals.

The multimodal perception module includes a selective attention submodule for information selection. Attention involves directing and concentrating mental activity toward specific stimuli, and selective attention functions to filter HCI-relevant information from vast amounts of irrelevant information [19]. Attention-guided selection of multimodal information includes bottom-up selection of salient information (e.g., flashing lights and loud alarms capture attention) and top-down selection of goal-relevant information (e.g., arrow signs in subway stations guide people to correct routes). Notably, the influence of the attention module extends beyond the perception module, also guiding components within the cognitive module.

(2) Cognitive Module

The cognitive module further processes information from the perception module, transforming it into executable operations while being regulated by attention. It comprises working memory, long-term memory, reasoning, and metacognitive submodules, with processing characteristics varying across individuals.

The working memory submodule is a capacity-limited information processing system for temporarily maintaining and storing information, serving as a platform connecting perception, long-term memory, and action output [20]. According to Baddeley's working memory model, working memory consists of three components: the central executive, visuospatial sketchpad, and phonological loop [20]. The central executive functions include shifting attention between tasks, selective attention and inhibition, updating and monitoring working memory contents, encoding temporal and spatial context in working memory, and planning subtasks to achieve goals. The visuospatial sketchpad primarily stores visual and spatial information, while the phonological loop specializes in storing speech-related information. Human working memory capacity is limited and relatively small, with individual differences [21]; computers can process substantially larger information volumes simultaneously.

The long-term memory submodule stores all knowledge and experiences from past learning and life events. Long-term memory constitutes an individual's "psychological past" and forms the basis for experience accumulation and cognitive development. It can be divided into declarative (explicit) and non-declarative (implicit) memory [22]. Declarative memory refers to knowledge individuals can consciously access, while non-declarative memory refers to knowledge inaccessible through conscious processes. Individuals possess large long-term memory capacities. Human long-term memory formation requires rehearsal and consolidation, taking considerable time. Computer long-term memory systems are more stable than human systems, as human memory retrieval can be subject to interference or modification [23].

Reasoning involves drawing conclusions from known information. Real-world individuals possess limited information and cognitive capacity, making decisions or judgments through heuristics such as representativeness heuristics (judging based on similarity to past experiences) and availability heuristics (deciding based on easily imagined or recalled events) [24]. Different individuals may draw different inferences from the same information.

Metacognition refers to humans' cognition of their own cognitive activities, comprising metacognitive knowledge, metacognitive experiences, and metacognitive monitoring [25]. Metacognitive knowledge consists of general knowledge about cognitive activities accumulated through practice, including knowledge about cognitive agents, tasks, and strategies. Metacognitive experiences are cognitive and emotional experiences occurring during cognitive activities, playing crucial roles—pleasant experiences can deepen cognitive activities, while frustrating experiences may terminate them. Metacognitive monitoring involves consciously monitoring and regulating ongoing cognitive activities. For example, successful recall requires overlap between retrieval information and encoding information, purposeful memory retrieval, attention to cues, and internal searching for desired memories.

(3) Action Module

The action module includes action matching, execution, monitoring, and correction. Action matching involves aligning cognitive module processing results with specific actions [26]. Actions possess hierarchical representations for specific goals [27]. The highest conceptual level involves abstract representations of action purposes derived from the cognitive module, with the action matching submodule selecting appropriate actions or action sequences to achieve goals. The lowest level translates these units into muscle movement patterns. Action execution involves the actual production of actions, with action sequences from the matching submodule sending commands to effectors, and the execution submodule completing corresponding actions. Different actions vary in learnability, operation time, and accuracy.

The action monitoring submodule oversees action execution through internal and external control. Internal control compares actual actions with planned actions to ensure correct execution, such as following a specific keypress sequence. External control involves exogenous cues guiding actions, such as pressing keys according to indicator lights. Action correction adjusts actions in real-time based on monitoring results to ensure goal achievement.

Individual actions are multimodal. In HCI, humans can act through traditional mouse and keyboard inputs or via pen, voice, eye movement, EMG, EEG signals (e.g., gaze-controlled web browsing tools [28], brain-computer interfaces using EEG signals [29]). Computers can also produce actions through images, sounds, mechanical movements, and other modalities. Computers must continuously expand their output channels—current interaction technologies primarily rely on vision and hearing with limited haptic feedback, far fewer than the ways humans perceive the world [30].

2.2 Shared Space

Humans and machines in HCI systems are not isolated individuals; human-machine hybrid intelligence requires collaboration to achieve specific goals. The shared interaction space describes the domain beyond individual processing spaces, divided into shared external space and shared internal space.

(1) Shared External Space

HCI systems encompass not only humans and computers but also their shared environment. The external environment influences all modules within individual processing spaces. For example, in hazardous environments, human hormone levels change, altering cognitive functions (such as attentional bias and working memory) and affecting action operations [31, 32]. Perceiving the external environment is crucial for individual survival and development. Situation awareness represents the core task for shared external space, with individuals selecting response patterns most appropriate to current contexts.

Environmental context changes interaction forms in human-human communication. For instance, in quiet libraries, communication tends toward text and gestures, while outdoor street scenes favor verbal communication. In HCI systems, consistency in situation awareness across individuals affects interaction naturalness. In smart speaker-user systems where voice is the primary interaction modality, both speaker and user should make consistent situational judgments based on the shared external environment, adjusting voice volume, vocabulary, and syntax accordingly.

(2) Shared Internal Space

The shared internal space refers to overlapping portions of perception, cognition, and action modules among individuals in HCI systems. Larger overlapping spaces enable more natural interactions. The shared internal space comprises shared experience modules and theory of mind working modules.

Shared Experience Module

The shared experience module includes two submodules: effective signal module and shared long-term memory module. It forms the foundation for sustained interaction behavior in HCI systems.

The effective signal module concerns the effectiveness of signal communication between individuals, calculated as the proportion of action information and individual state information emitted by one individual's action execution module that is received by other individuals' multimodal perception modules. True states or intentions are often expressed through multiple pathways. For example, individuals can recognize emotions through facial expressions, but identical facial expressions with different body postures convey different emotions [33]. Individuals must receive multichannel information from others, as natural HCI relies on complete information understanding between individuals.

The shared long-term memory module refers to long-term memory content shared among all individuals in HCI systems. Common knowledge backgrounds enable more natural human-human interactions. For example, experts from the same domain can engage in more natural and harmonious complex interactions due to similar knowledge backgrounds.

Theory of Mind Module

Theory of mind refers to the complex metacognitive understanding of mental states, forming the basis for individuals to explain and predict their own and others' behaviors [34]. Key components include understanding others' intentions, joint attention, and advanced theory of mind [35]. Understanding others' intentions is a crucial HCI goal—namely, narrowing the evaluation gap. Joint attention is a vital human characteristic; when others shift their gaze, humans comprehend that individuals are attending to something and shift attention to the same target. Learning robots with joint attention modules can promote children's learning. Advanced theory of mind refers to complex knowledge bases about individuals acquired during HCI processes, which help individuals understand others' behaviors and states.

The theory of mind module guides work across all modules in individual processing spaces. It can adjust an individual's own action execution content. For example, in human-robot collaboration scenarios where robots guide users to complete target construction tasks, robots with theory of mind modules describe object locations and operation directions from the user's perspective rather than their own. Individuals are not always perfectly accurate in perception, cognition, and action matching/execution. Errors in natural HCI are shared experiences among all system individuals, and many errors are immediately corrected under theory of mind guidance. For instance, when a user's operation is misinterpreted by a hardware system, the user immediately understands the hardware system's internal knowledge architecture or detects their own violation of shared space principles in action standards, promptly changing actions and executing new operations.

The shared internal space describes connections between individuals and machines, continuously updating through these connections. The shared experience module accumulates throughout interaction processes. In HCI, individuals' common interaction experiences are added to the shared long-term memory module. Simultaneously, under theory of mind guidance, individuals gradually form complex knowledge bases about other individuals' characteristics. For example, in sketch-based interaction, individuals gradually learn and understand the meanings of specific symbols used by others, storing such symbol-meaning associations in the shared long-term memory module. Through ongoing sketch interaction, individuals also learn unique characteristics of others, such as distinctive handwriting features.

3 Conclusion and Outlook

This paper proposes a Shared Mental Model based on classical cognitive models such as the Model Human Processor and the characteristics of natural HCI. The model inherits and develops the traditional three modules of perception, cognition, and action, and innovatively introduces a shared space module. Natural HCI requires humans and machines to possess maximally large shared spaces, specifically extensive shared experiences and sophisticated theory of mind. This model can provide evaluation frameworks and optimization guidelines for interaction naturalness in various emerging HCI paradigms, including multi-user and multi-device systems.

The Shared Mental Model attempts to provide new theoretical guidance and development recommendations for natural interaction, though certain limitations remain and require further refinement. First, the model is currently theoretical; substantial data support is needed to develop it into a computational model. The model also requires validation with real user data integrated with specific interaction forms and devices. Second, while the theoretical model provides a basic framework for natural interaction, specific content within each module will vary in emphasis across different interactive products. The model's practical application value must be examined across various interaction scenarios. Finally, although natural interaction represents the future development trend, controversies remain regarding definitions and evaluation systems of naturalness, as mentioned at the outset. Achieving theoretical unification will require collaborative efforts from scholars across multiple disciplines.

References

[1] Grandhi S A, Joue G, Mittelberg I. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces[C]//Proceedings of the SIGCHI conference on human factors in computing systems. 2011: 821-824.
[2] Norman D A. Natural user interfaces are not natural [J]. interactions, 2010, 17(3): 6-10.
[3] Jain J, Lund A, Wixon D. The future of natural user interfaces[M]//CHI'11 Extended Abstracts on Human Factors in Computing Systems. 2011: 211-214.
[4] Wigdor D, Wixon D. Brave NUI world: designing natural user interfaces for touch and gesture [J]. ACM SIGSOFT Software Engineering Notes, 2011, 36(6): 29-30.
[5] Norman D A, Nielsen J. Gestural interfaces: a step backward in usability [J]. interactions, 2010, 17(5): 46-9.

[6] 史元春. 自然人机交互 [J]. 中国计算机学会通讯, 2018, 14(5): 8-10.

[7] Card S K. The psychology of human-computer interaction [M]. Crc Press, 2018.

[8] Anderson J R, Bothell D, Byrne M D, et al. An integrated theory of the mind [J]. Psychological Review, 2004, 111(4): 1036.

[9] Liu Y, Feyen R, Tsimhoni O. Queueing Network-Model Human Processor (QN-MHP) A computational architecture for multitask performance in human-machine systems [J]. ACM Transactions on Computer-Human Interaction (TOCHI), 2006, 13(1): 37-70.

[10] 王晖, 戴国忠, 禤宇明, 等. 虚拟现实人机交互的 eGOMS 模型研究 [J]. 华中科技大学学报（自然科学版）2005 (z1): 322-325.

[11] 刘烨, 汪亚珉, 卞玉龙, 等. 面向智能时代的人机合作心理模型 [J]. 中国科学: 信息科学, 2018, 48(4):

[12] Wu C, Liu Y. Queuing network modeling of driver workload and performance [J]. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(3): 528-37.

[13] 刘伟. 人机混合智能: 新一代智能系统的发展趋势 [J]. 上海师范大学学报 (哲学社会科学版),2023,52(01).

[14] 张警吁, 张亮. 自然交互的认知机理与心理模型 [J]. 中国计算机学会通讯, 2018, 18(5): 30-5.

[15] Spence C. Multisensory perception [J]. Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 2018, 2(1-56.

[16] Lalanne C, Lorenceau J. Crossmodal integration for perception and action [J]. Journal of Physiology-Paris, 2004, 98(1-3): 265-79.

[17] Mcgurk H, Macdonald J. Hearing lips and seeing voices [J]. Nature, 1976, 264(5588): 746-8.

[18] Metatla O, Maggioni E, Cullen C, et al. " Like Popcorn" Crossmodal Correspondences Between Scents, 3D Shapes and Emotions in Children; proceedings of the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, F, 2019 [C].

[19] Sanders L D, Astheimer L B. Temporally selective attention modulates early perceptual processing: Event-related potential evidence [J]. Perception & psychophysics, 2008, 70(4): 732-42.

[20] Baddeley A D, Hitch G. Working memory [M]. Psychology of learning and motivation. Elsevier. 1974:

[21] Unsworth N, Engle R W. The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory [J]. Psychological review, 2007, 114(1): 104.

[22] Rutishauser U. Testing models of human declarative memory at the single-neuron level [J]. Trends in cognitive sciences, 2019, 23(6): 510-24.

[23] Sinclair A H, Barense M D. Prediction error and memory reactivation: how incomplete reminders drive reconsolidation [J]. Trends in neurosciences, 2019, 42(10): 727-39.

[24] Artinger F, Petersen M, Gigerenzer G, et al. Heuristics as adaptive decision strategies in management [J]. Journal of Organizational Behavior, 2015, 36(S1): S33-S52.

[25] Norman E, Pfuhl G, Sæle R G, et al. Metacognition in psychology [J]. Review of General Psychology, 2019, 23(4): 403-24.
[26] Pezzulo G, Cisek P. Navigating the affordance landscape: feedback control as a process model of behavior and cognition [J]. Trends in cognitive sciences, 2016, 20(6): 414-24.
[27] Mackay D M. Towards an information-flow model of human behaviour [J]. British Journal of Psychology, 1956, 47(1): 30-43.
[28] Menges R, Kumar C, Müller D, et al. Gazetheweb: A gaze-controlled web browser[C]//Proceedings of the 14th International Web for All Conference. 2017: 1-2.
[29] Aricò P, Borghini G, Di Flumeri G, et al. Passive BCI beyond the lab: current trends and future directions [J]. Physiological measurement, 2018, 39(8): 08TR2.
[30] Obrist M, Ranasinghe N, Spence C. Multisensory human–computer interaction [J]. International Journal of Human-Computer Studies, 2017, 107(
[31] Jiang C, Buchanan T W, Yao Z, et al. Acute psychological stress disrupts attentional bias to threat-related stimuli [J]. Scientific reports, 2017, 7(1): 1-11.
[32] Lin L, Zhang J, Wang P, et al. Perceived control moderates the impact of academic stress on the attention process of working memory in male college students [J]. Stress, 2020, 23(3): 256-64.
[33] Aviezer H, Hassin R R, Ryan J, et al. Angry, disgusted, or afraid? Studies on the malleability of emotion perception [J]. Psychological science, 2008, 19(7): 724-32.
[34] Gallagher H L, Frith C D. Functional imaging of 'theory of mind' [J]. Trends in cognitive sciences, 2003, 7(2): 77-83.
[35] Baron-Cohen S. Mindblindness: An essay on autism and theory of mind [M]. MIT press, 1997.

(Corresponding author: Ma Cuixia, E-mail: cuixia@iscas.ac.cn)

Author Contributions:
Zhang Liang, Wang Xiaoyu, Liu Fang, Ma Cuixia: Theoretical construction;
Zhang Liang, Wang Xiaoyu: Draft preparation;
Zhang Liang, Ma Cuixia, Wang Hongan: Revision of final manuscript.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.