Many Meanings of Intentionality: A Brief Disambiguation for Computational Creativity

Simo Linkola and **Iikka Hauhio**
Department of Computer Science
University of Helsinki, Finland
simo.linkola@helsinki.fi, iikka.hauhio@helsinki.fi

Abstract

Computational creativity is a multi-disciplinary field with authors from diverse backgrounds. This raises a threat of misunderstanding when the authors from different backgrounds use the same words with different meanings. We elaborate on the two main meanings of the word "intentionality" found in the computational creativity literature: aboutness and goal directness. Both of the meanings are prominent concepts in computational creativity, but they relate to very different ideas, perspectives and contexts. Aboutness, in philosophy, is quality of the mind to be about somethings outside of it, while goal directness, when interpreted from engineering perspective, is the system's property to be able to produce outputs that are aligned with its goals. We briefly explain both of the meanings, highlight their related concepts, and provide a discussion of how these two interpretations of the same word are related.

Introduction

Creativity research entails many related concepts and terms that are argued to be relevant for creativity. One of these terms is "intentionality", which has even been considered as a necessary requirement for computational creativity (CC) from a process perspective (Ventura 2016; 2017) and appended to the artefact (or concept) requirements of the standard definition of creativity including novelty and value (Boden 2004; Runco and Jaeger 2012). Unfortunately, the word "intentionality" has multiple homonymous meanings which are seemingly completely separate. This can cause unnecessary confusion in the field when researchers using the word – or referring to others using the word – do not specify exactly which meaning they intend. In this short paper, we briefly explain (some of) these different meanings and their implications, and suggest that the field, in general, takes care when using the word – or select more fitting terms which have no similar tendency to cause confusion.

Computational Creativity is a multi-disciplinary field with contributing members from different fields such as computer science, psychology, fine arts and philosophy. Because of this multi-disciplinary approach, the field is able to incorporate many perspectives to the multi-faceted phenomenon of creativity, which has been a discussion subject within the field in its own right. For example, Pérez Y Pérez (2018) describes a CC-continuum which spans from engineering-mathematical approach to cognitive-social approach, where the ends of the continuum follow different paradigms and may have different research methods as well as end goals. This produces internal tension, typical to many multi-disciplinary fields, and the field has to constantly battle with misinterpretations of the conveyed messages as the vocabulary between the members of the field may vary greatly.

In this paper, we focus on the meaning of a single word, "intentionality", because of its central status within the field of computational creativity. In philosophy, also adopted to cognitive science, intentionality means "aboutness", the ability of the mental states to refer to objects outside the mind (Schlicht and Starzak 2021). In other contexts, intentionality may refer simply to the property of having intentions, goals, or objectives (Ventura 2016; 2017). At a first look, these meanings have nothing to do with each other but sharing the same name, yet the word is used without referring to its explicit meaning in the existing computational creativity literature (Colton and Ventura 2014; Grace and Maher 2015; Varshney 2020; Sewell, Christiansen, and Bodily 2020).

The rest of the paper is organised as follows. Next, we provide a brief disambiguation of the two main meanings of the word "intentionality" especially relevant for the field of computational creativity. Then, we address how the different meanings may possibly relate to each other. We end the paper with conclusions.

Interpretations of Intentionality

In this section, we review the different meanings of the word "intentionality" relevant for computational creativity. We begin with the philosophical concept of *aboutness* and follow with more engineering-oriented *goal directness*.

Aboutness

In philosophy, intentionality refers to the quality of mental states to be "about" something or "directed to" something (Schlicht and Starzak 2021), popularised by Franz Brentano, who argues that it is what separates mental states from physical states (Brentano 1874). While intentionality is a quality of mental states, it applies to words and symbols as well if they are being processed by a mind. That is, an intentional mind can produce meaningful words and symbols, and as well interpret words and symbols as having a meaning.

Intentionality of computer programs is a debated topic. Searle (1980) argues with his famous Chinese Room thought experiment that computers – and computer programs – are not intentional, and intentionality is exclusively a feature of biological brains and "equivalent" systems. Following Searle's thought experiment, Harnad (1990) formalised the question of how symbolic computation systems could become intentional into the so-called *symbol grounding problem*: how can symbols have meaning, if they are only defined in terms of other symbols in the computation system? Harnad argues that in order for the system to be intentional (i.e., its symbols be grounded), they must have intrinsic, not extrinsic meaning. He calls extrinsic meanings *parasitic*, since they rely on outside observers for interpretation. That is, if the program is not intentional, its states are meaningless by themselves. Conversely, the states of an intentional program have self-standing meanings.

Intentionality is an overarching property of all mental states, and is thus not only applied to communication. However, in the context of generative artificial intelligence, the focus is most often in the inputs and outputs of the program. When analysing intentionality pertaining to the output, the term *communicative intent* (Bender and Koller 2020) or *authorial intent* (Barten 1967) is often used to refer to the meaning of the output as "intended" by its producer. This term has some overlap with the term "goal" as in "goal directness": the communication is typically done for a purpose, and thus the communicative intent is often directly related to a goal of a program. However, not all communicative intents are necessarily tied to a goal: for example, a person might accidentally say something that is against their goals, in which case the words still have meaning, but they are not aligned with the goals of the person. Another example might be a person speaking their stream-of-consciousness, saying what comes to their mind. In both of these cases, the speech act as a whole might still have a goal, but the individual parts of the act might not.

Intentionality can be seen as a crucial part of both creativity and cognition in general. It enables understanding (i.e., retrieving meaning of) mental states, including one's own processes, goals, memories, and so on, but also interpreting and determining value of texts and other artefacts. If a program is not intentional, its creativity and the value of its works are parasitic, dependent on outside observers.

However, not all consider intentionality important. Most famously, Roland Barten argues in his essay "The Death of the Author" that authorial intent should play no role in the interpretation of a text (Barten 1967). According to his view, literary works should be regarded as eternal objects, discovered rather than created, that have "no origin but the language itself". Under this kind of viewpoint, the communicative *intent* of the author no longer matters, but merely the communicative *function* of the words, i.e., how they are actually interpreted: if it produces something novel and valuable for the observer, it doesn't matter how it did it.

Goal Directness

In the field of computational creativity, Ventura (2016; 2017) defines intentionality as

> the fact of being deliberative or purposive; that is, the output of the system is the result of the system having a goal or objective – the system's product is correlated with its process.

Ventura (2016; 2017) does not provide an explanation for his definition to let us better understand the influences behind it; we interpret this to mean that he assumes the definition to be general enough – perhaps in line with the everyday usage of the word – that it can be accepted without explicit references. Thus, next, we provide a brief disambiguation of the definition to explain how we interpret the definition and what other possible – though mostly improbable – interpretations there could be.

The above textbook definition conflates the property of (1) being deliberative or purposive, with the notions that (2) the output of the system is the result of the system having a goal and (3) the output of the system correlates with the process of the system. All three statements of the definition are used to illustrate what intentionality in general terms means without having to construct a more formal definition, yet each of them alone may be interpreted to mean something subtly different.

Being deliberative or purposive may refer to a human (or animal) capability. That is, this statement alone can be interpreted to imply that the entity with this property has similar mental state quality as "aboutness". On the other hand, it may imply that there are other properties, such as the next two statements which are used in the definition to clarify what the first statement means.

The output of the system is a result of the system having a goal can be interpreted at least in two different ways: (1) the system has a goal to produce (certain kind of) outputs in general or (2) the production of a single output is affected by a goal. The main difference between these two interpretations is the time scale: on (1) the focus is on the general functionality and goals of the system while on (2) the focus is on a single artefact production process. The second interpretation can be further elaborated as (2a) the production of the output began because of a particular goal or (2b) the output aims to represent a particular goal of the system.

The system's product is correlated with its process can be interpreted in as many ways as an output can be correlated with the process it was produced by. One straightforward interpretation is that the process producing the output varies – somehow meaningfully or within reason – based on what the outputs aim to represent. That is, the process varies based on the goals for this particular output.

The most likely interpretation of the last two statements forms a description of a system producing outputs which aim to represent particular goals (2b, above), and how the system produces the particular outputs is affected by the goals aimed to be represented or fulfilled by the outputs. In other words, *the goals of the system (for particular outputs) affect the production process and, thus, the outputs*.

The above interpretation of Ventura's intentionality definition is in line with the concept of *goal directness*, a property where the behaviour of the system is aimed towards a goal or a completion of a task. This interpretation does not state anything about how this behaviour is achieved, yet if the first statement of Ventura's definition is also taken into account, it may be implicitly assumed that the system has human characteristics such as "aboutness".

Unfortunately, goal directness is another term subject to misinterpretation as its origins are in psychology (Frese and Sabini 2021). However, the term is frequently used in the context of artificial intelligence and intelligent agents loosely in the same way as the last two statements of Ventura's intentionality definition. Goal-based and utility-based agents (Russell and Norvig 2010) both have the capability of goal-directed behaviour, while learning agents can do so in adaptive deployment environments.

Overall, goal directness is associated with many other terms and concepts, e.g., in philosophy, artificial intelligence, and software engineering.

First, in software engineering, self-adaptive and self-aware systems (Kounev et al. 2017) aim to account for more appropriately exhibited goal-directed behaviour by allowing the system to change how it operates towards its goals based on the observed context and the system state. Linkola et al. (2017) elaborate on the concept of self-awareness in the context of software architectures for artificial, creative systems. By their argumentation, goals are one of the most notable aspects for the creative systems to be self-aware of.

Second, the question of from where the goal-directed behaviour originates, i.e., what *motivates* it and who selects the goals, is interesting. On the philosophical side of computational creativity, Guckelsberger, Salge, and Colton (2017) have studied the notion of *why* a creative system does what it does, arriving to the conclusion that, in the end, it is nearly always the programmer who decided the goals. Furthermore, motivation of intelligent agents in general has been discussed, e.g., in the context of reinforcement learning (Schmidhuber 2010; Barto 2013).

Third, the concept of creative autonomy (Jennings 2010) is related to not only goal-directed behaviour but also to motivation and self-awareness. To have creative autonomy, the system must fulfil three requirements: autonomous evaluation, autonomous change and non-randomness. Autonomous evaluation states that the system must be able to evaluate its own outputs. Thus, it directly relates to goal directness, as being able to evaluate what the system itself did, in many cases, provides ways for the system to reach its goals better. Autonomous change states that the system must be able to change its own evaluation standards, which relates to the meta-level discussion of goals and their adjustments, a prominent focus in self-adaptive and self-aware systems. Lastly, the non-randomness states that the system's evaluation or change of evaluation standards is not purely random. This statement relates to the process part of the goal-directed behaviour. While it does not state which processes should be used or how much non-randomness there should be, it is argued that concepts of self-adaptive and self-aware systems as well as motivation can help in satisfying this requirement.

Bridging the Gap

At first sight, when comparing the philosophical concept of aboutness and the engineering interpretation of goal directness, it may seem that these two concepts have little in common. However, their relationship becomes more evident if we take a look of the terms related to goal directness in the field of psychology.

In early behaviourism (Frese and Sabini 2021), to avoid long-winded discussion, goal-directed behaviour was associated with the problems of *teleology*, the idea that the future acts upon the past, and *prevision*, a plausible account of the anticipation of a goal. Prevision can be further explained, e.g., with *representations* by which an organism can evaluate the results of their behaviour, but the dichotomy of having a purpose and how to act on that purpose is not resolved. Nonetheless, an organism must have aboutness in order to have representations of the results of the behaviour to associate the mental representations with the real world.

In more modern psychology (Frese and Sabini 2021), the notion of the negative feedback between the goal state and the current state avoids the teleological conundrum; the goals of the behaviour can affect the behaviour which is aimed at fulfilling those particular goals. This feedback loop, as an abstract notion, does not require mental states to have any particular quality, and it can be used to provide structure to the behaviour, e.g., long-term planning. The very idea of feedback loop is also a basic building block, in various forms, in modern intelligent agents (Russell and Norvig 2010; Barto 2013), and thus the usage of the term goal-directed behaviour is apt for intelligent agents from this perspective.

However, there are still some nuances that are not captured with the above elaboration of goal-directed behaviour when relating it to aboutness. Aboutness is a quality of the mental states; a mental state is about something and that something can be a cat, a building plan, the objective function of an AI program, or the communicative intent we assume another person to have. That is, aboutness is an overarching quality of the mind with plethora of application targets depending on the context and priming. Goal directness, on the other hand, does not specify whether the system can interpret a single type of goal or multiple types of goal, and how abstract these goal types are. That is, to even begin to argue that an artificial system has some notion of human aboutness in itself, it should be able to reason about vast number of different concepts on different abstraction levels – potentially changing the reasoning domain or process during the procedure. This kind of behaviour is not often covered or measured by the concept of goal directness in practical artificial intelligence or computational creativity.

Aboutness as a quality of mind that can manifest reflection for different phenomena, not only for the behaviour of oneself, brings about further differences. For example, aboutness can be present in elaborating the communicative intent the other person has when conveying a message. This kind of reflection, which is not directly about the environment and the agent's goals, is not always covered by AI techniques which can be deemed to satisfy some level of goal directness. While the AI field also tackles these problems, the fact that they require different solutions implies that the aboutness, as understood in philosophy, is a more general phenomenon than what AI techniques currently cover.

In artificial systems, both of the above differences of functionality provided by aboutness with respect to functionality assumed by goal directness, multiple application targets and multiple abstraction levels, can be tackled to some extent with meta-reasoning and other meta-level approaches such as self-adaptive and self-aware systems. However, the main philosophical debate still exists: whether a computational system can have aboutness as its own quality or not.

Conclusions

In this paper we have discussed the meaning of the word "intentionality" in the context of computational creativity. We have explained the two main interpretations, aboutness and goal directness, and provided a brief analysis on how these interpretations can be aligned with respect to each other. The main point of this disambiguation was to show how these interpretations are different and what concepts or qualities of the interpretations are related.

Overall, we propound authors of computational creativity papers to be aware of these different interpretations and, should they use the term in their papers, clearly state which interpretation they are referring to. Using the same word with different interpretations – some of which are conceptually more challenging to replicate in machines – may cause dilution of the multi-faceted and nuanced philosophical concepts into luke-warm engineering solutions.

Acknowledgments

SL is funded by Academy of Finland grant 328729 (CACDAR). IH is funded by the doctoral programme in Computer Science at the University of Helsinki. The authors wish to thank Anna-Mari Rusanen for her advice on the topic of intentionality in philosophy.

References

Barten, R. 1967. The Death of the Author. *Aspen*.

Barto, A. G. 2013. Intrinsic motivation and reinforcement learning. *Intrinsically motivated learning in natural and artificial systems* 17–47.

Bender, E. M., and Koller, A. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 5185–5198.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. London, UK: Routledge, 2nd edition.

Brentano, F. 1874. Psychology from an empirical standpoint (trans. L. MacAlister). *New York 1973*.

Colton, S., and Ventura, D. 2014. You can't know my mind: A festival of computational creativity. In *Proceedings of the Fifth International Conference on Computational Creativity*, 351–354.

Frese, M., and Sabini, J. 2021. Goal directed behavior: The concept of action in psychology.

Grace, K., and Maher, M. L. 2015. Specific curiosity as a cause and consequence of transformational creativity. In *Proceedings of the Sixth International Conference on Computational Creativity*, 260–267.

Guckelsberger, C.; Salge, C.; and Colton, S. 2017. Addressing the" why?" in computational creativity: A non-anthropocentric, minimal model of intentional creative agency.

Harnad, S. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1-3):335–346.

Jennings, K. E. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 20(4):489–501.

Kounev, S.; Kephart, J.; Milenkoski, A.; and Zhu, X. 2017. *Self-Aware Computing Systems*. Springer International Publishing.

Linkola, S.; Kantosalo, A.; Männistö, T.; and Toivonen, H. 2017. Aspects of self-awareness: An anatomy of metacreative systems. In *Proceedings of the Eight International Conference on Computational Creativity*, 189–196. Atlanta, Georgia, USA: Georgia Institute of Technology.

Pérez Y Pérez, R. 2018. The computational creativity continuum. In *Proceedings of the Ninth International Conference on Computational Creativity*, 177–184. Salamanca, Spain: Association for Computational Creativity.

Runco, M. A., and Jaeger, G. J. 2012. The standard definition of creativity. *Creativity research journal* 24(1):92–96.

Russell, S. J., and Norvig, P. 2010. *Artificial intelligence: A modern approach*. Pearson Education, Inc.

Schlicht, T., and Starzak, T. 2021. Prospects of enactivist approaches to intentionality and cognition. *Synthese* 198(Suppl 1):89–113.

Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* 2(3):230–247.

Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and brain sciences* 3(3):417–424.

Sewell, A.; Christiansen, A.; and Bodily, P. M. 2020. Creative constellation generation: A system description. In *Proceedings of the 11th International Conference on Computational Creativity*, 496–499.

Varshney, L. R. 2020. Limits theorems for creativity with intentionality. In *Proceedings of the 11th International Conference on Computational Creativity*, 390–393.

Ventura, D. 2016. Mere generation: Essential barometer or dated concept? In *Proceedings of the Seventh International Conference on Computational Creativity*, 17–24. Paris, France: Sony CSL.

Ventura, D. 2017. How to build a CC system. In *Eighth International Conference on Computational Creativity*, 253–260. Atlanta, GA, USA: Association for Computational Creativity.