Croatian Journal of Philosophy
Vol. XXII, No. 66, 2022
https://doi.org/10.52685/cjp.22.66.8
Received: July 30, 2022
Accepted: September 17, 2022

Purposiveness of Human Behavior. Integrating Behaviorist and Cognitivist Processes/Models

CRISTIANO CASTELFRANCHI
National Research Council, Rome, Italy

We try not just to reconcile but to "integrate" Cognitivism and Behaviorism by a theory of different forms of purposiveness in behavior and mind. This also implies a criticism of the Dual System theory and a claim on the strong interaction and integration of Sist1 (automatic) and Sist2 (deliberative), based on reasons, preferences, and decisions. We present a theory of different kinds of teleology. Mere "functions" of the behavior: finalism not represented in the mind of the agent, not "regulating" the behavior. Two kinds of teleological mental representations: true "Goals" in control-theory, cybernetic view, with "goal-driven" behavior (intentional action); vs. Expectations in Anticipatory Classifiers: a reactive but anticipatory device, explaining the "instrumental" (finalistic) nature of Skinner's reinforcement learning. We present different kinds of Goals and goal processing and on this ground the theory of what "intentions" are. On such basis, we can discuss Kathy Wilkes's hint about the necessarily linguistic formulation of "intentions"; with the hypothesis that her intuition is not correct for any kind on "intention" which may be represented in sensory-motor format, but correct for "volition" and our will-strength for socially influencing ourselves.

Keywords: Teleology; goal theory; intentions; behaviorism; dual System.

1. *Premise: Claims and Moves**

The *claims* are the following ones:

It is time—also thanks to the pressure due to the neuro-foundation of psychological models—to reconcile Cognitivism with Behaviorism (two philosophical and historical enemies). Not just to reconcile but to "integrate" them, by not simply explaining coexistence of postulated mechanisms but their systemic interaction and interference. This attempt will in part overlap with a reunification of System 1 and System 2 postulated in the "Dual System" view of the mind.

Main moves necessary for this integrated theory in fact are:

– A critical revision of "dual process" theory:[1]

(i) It assembles as a unified "process" (automatic, fast, associative, holistic) several very different mechanisms; or just opposes "affect" and "reason" (Loewenstein and O'Donoghue 2004)

(ii) These ("multiple" not "dual") processes do not just compete and prevail one on the other, but interact and cooperate (for example, in the complex and hybrid "value" of a goal, both belief-based, reasoned, and just "felt" ("somatic markers", etc.).

– Making formally clear the fundamental distinction between the *two kinds of finality*, of "goal", impinging on animal behavior: *mental goals* (based on control theory models), vs. *external goals, mere "functions"* (based on selection processes). A frequent mistake of psychologists (Castelfranchi,1999) is to interpret any clear purposefulness of human behavior in terms of conscious or unconscious intentions in the mind of the individual (Bargh et al. 2001).

– In this frame, we need—as said—a more "representational view" of conditioning.[2] However, in the "mentally represented" teleological devices we will distinguish true "Goals" from Expected Results reinforcing and explaining that conduct. It is crucial to make clear the difference between these *two kinds of anticipatory representation* governing the action. And modeling on such basis the "instrumental" (finalistic) nature of Skinner's conditioning.

– Modeling the layered *integration* of reactive/automatic devices and of intentional and reasoned actions; for example, by implementing higher level deliberated action in underlying automatic classifiers.

One should also try to:

– Explain how conditioning, reinforcement learning (both Pavlovian and instrumental), act also on symbolic "mental representations" pos-

[*] This is more a palimpsest of a work in progress than a balanced paper. It contains a vision and some basic claims; a schema of the main moves that should be done; and exploration of a few specific issues including an homage to Kathy Wilkes' intuitions.

[1] Nowadays very popular. Literature is very broad and with different positions (Caccioppo, Kahneman, Sloman).

[2] And putting aside some really reductive proposal of behaviorism, like the reduction of guilt feeling to worry for punishment!

tulated by Cognitivism (beliefs, expectations, goals...), and interfere (not only compete) with the high-level cognitive processes.
– To discuss the notion of "reward" and its function, and to put aside "hedonism" (pleasure) as the unifying motivation.

Let us be a bit more analytical on some of those issues. At the end, on the basis of theory and modeling of "intention", we will discuss an interesting thesis of Kathy Wilkes, as an homage to her deep thinking.

2. *The anticipatory nature of the mind: two devices*

It is very important to understand the anticipatory nature/origin of mind (and the more general "augmented reality" function of the brain) and the creation of "endogenous" representations/worlds: not output of current perception input, but self-generated by memory activation, generative recombination, imagination and simulation. A fictional world where to act, learn, solve problems.

However, we have to distinguish two very different anticipatory devices: Anticipatory Classifiers (ACs) (bottom-up, responsive) versus true goals (control theory, top-down) (Pezzulo et al. 2008). In both cases, there are "expectations" but with different roles: in ACs just reinforcement function; in Goals cybernetic set-points, monitoring and adjusting, (sub)planning. In both cases, there is "failure" (frustration) or "success."

ACs are very important for contrasting a primitive behaviorist, conditioning-based explanations of some behaviors just in terms of S-R, Condition-Action ("production rules" or Classifiers) models of reinforcement learning.

However, we also have to be reminded that there is another kind of finalism in animal and human conduct, not represented at all: mere "functions" of that behavior (or feature).

2.1. *Mere "functions" as not mental/represented goals*

As already said there are *two kinds of teleology*: (i) mentally represented (and eventually intended) or *psychological goals* that regulate our conduct; and (ii) non-mental goals, just emergent and self-organizing "functions" (social or biological) impinging on our individual and collective behaviors. Let us use the term "goal" just for the internal control system, the mentally represented objective; and the term "function" for the external selecting finality of a feature or a behavior (Conte 1995; Castelfranchi 2001).

Behavioral functions are simply effects of behavior, which give a positive feedback on it, reinforce or select it, and reproduce it. *Functional effects*, usually unintended (desirable or even undesirable) and not understood, but such that they have feedback and select that behavior or entity.
– *Selective/evolutionary "functions"* of behaviors or features (not only of behaviors; also the features of living being have a function: an adaptive effect)

There also are:

– *Technical functions:* objects also have finalisms: they are "made *for*" and "used *for*": function of the object / tool.

Also in cognitive intentional Agent there can also be merely "function" governed conducts: effects of behavior which go beyond the intended effects but which can successfully be reproduced because they reinforce the agent's beliefs and goals that give rise to that behavior.

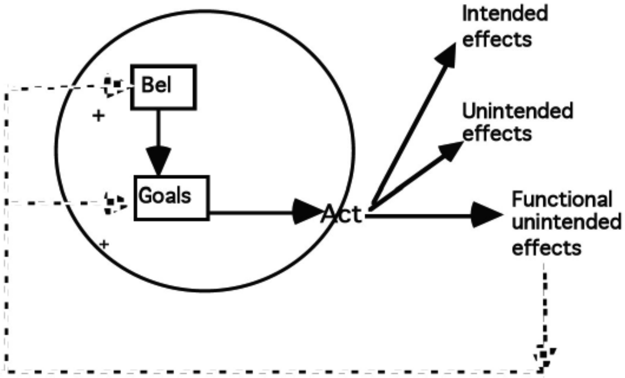

Fig.1 Functional unintended effects

2.2. *Teleology in Dual Processing System 1*

Do we intend all the goals/finalities of our behavior? We do not "intend" all that we "pursue" ("functions"). Are all the expected positive results, the achieved goals "intended" results? No, we do not "intend" all we expect. As presented in this frame, we need a more "representational view" of conditioning first of all by making clear the difference between two kinds of *anticipatory representation* governing the action: true "goals" for goal-directed action vs. "anticipatory classifiers"—as special kind of "classifiers".[3]

The format of *Anticipatory Classifiers* is: C → A + Exp

Matching Condition activates an Action + Expectation (anticipated results).[4]

Similar to intentions but not intentions: not a "goal-driven" behavior whose model is TOTE model of Miller, Galanter, Pribram (1968) characterized by a top-down processing (from the goal to the action) not a bottom-up process:

[3] They are "Classifiers" (Cond ⇒ Action, S-R like) but they are based on *Anticipatory Representation*, on Expectations. Condition → Action + Exp. And their reinforcement is *due to the confirmation of the expected result* (Exp) (Pezzulo et. al. 2008).

[4] Moreover, Exp, or anticipated representation of perceptual nature, is *an expected sensation that* determines the "success" or failure of the act. Sensation that might also be *proprioceptive* or *enteroceptive*, that is, about a bodily state: a "feeling'.

First comparing the GOAL (starting point) against the World, and then (in case of mismatch) searching for/activating an action.

This kind of *Proto-Goals*[5] (Exp in ACs) and *proto-intentional conducts* are important in human agents for several reasons/functions:

- evolutionary and developmental stages;
- coexistence of different *teleonomic* mechanisms (not simple S-R) that govern and contend for behavior;
- Routine and automatic components of conduct; also of the intentional conduct;
- For explanation of—in our view—"Instrumental or Operant" conditioning and learning (Skinner), and why it is *seemingly intentional;*
- Probably also for explaining the "reinforcement learning" component of *neurotic persistence*, and its *circularity* in particular, when combined with the idea of sensorial and especially entero-ceptive expectations (feelings), sensations from/about my own body ex."relief" as a reinforcing—non realized—experience/feeling (ex. social anxiety; avoidance).

Moreover this mechanism and this anticipated representation and expectation is not necessarily conscious. The subject can be unaware of it, and this kind of primitive "control" can be merely "automatic" (like using the brakes and expecting the car to slow down) (Castelfranchi 2001).

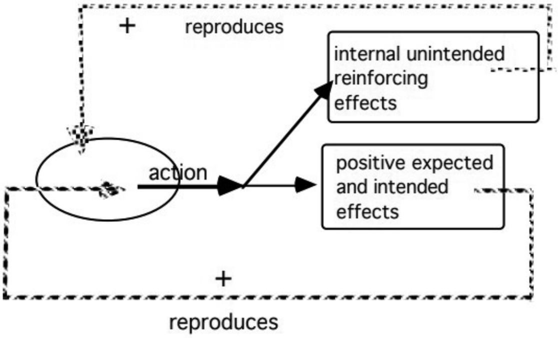
+
reproduces
internal unintended reinforcing effects
action
positive expected and intended effects
+
reproduces

Fig. 2 Unintended effects reinforcing the conduct

In our view, for example, ACs are crucial for explaining the "reinforcement learning" component of *neurotic persistence*, and its *circularity*. We wonder (but I am not a clinical psychologist!) if this dynamics is underlying "akrasia" experience in general. When I act in conflict with my best preference, what I think of would be better to do. We do not think that such a conflict and scission is just a conflict between affective im-

[5] "Proto" because they are similar to but not true goals, but also because reasonably they were the first form of mentally represented results, anticipated, and finalizing the conduct; before true goals and intentional actions.

pulses versus reasoned planning (like in Lowenstein version of "dual processing" (Loewenstein and O'Donoghue 2004)). Nor do we think—see below—that this is the result of a double reasoned decision process were there are *consciously* calculated advantages but also (prevailing) *unconsciously* calculated "secondary" outcomes with greater utility. Are neurotics perfectly *crypto-rational decision makers*? We guess that it is a matter of a conflict between a merely conditioned activated conduct vs. an intention-driven attempt.

How many human conducts are read as strictly goal-driven (intentional, preferred) while they are just conditioned?

2.3. *"Secondary advantage"*

In our view (Castelfranchi 1998, 1999) "secondary advantage" exists and operates, but it is not a "calculated" advantage we put in our reasoned decision, and we "rationally" *decide* for it but unconsciously (against what we consciously believe to prefer and would like to do). We are not rational but *unconscious decision makers*. The behavioral output is not the outcome of a reasoned evaluation of pros and cons; we do not choose what we consider better for us. The underlying model is a different one; it is a DUAL processing model, where two systems (the automatic, nonintentional reinforcement basic one, and unconscious and the deliberate one) compete with one the other, and the system based on "instrumental" reinforcement learning and on anticipation (but not "intention"!) of the reward can win, and we do something different (perhaps even do not really understanding "why") from what we would reasonably prefer. And we perhaps find some post-hoc and ad hoc explanations (reasons) of our choice, not necessarily the right ones! We expect a reward and act "in order" to obtain such (internal) reward (pleasure, pain avoidance, relax, stop anxiety…) but such an expectation is not our "goal" in control theory and psychological sense. It is just the Exp of an anticipatory classifier, maintained/reinforced by its activation, execution, and success/confirmation of the result. We are forced by such reactive and sensation-based but prospective device. And we can in fact also feel "without control and real decision," acting against our good and intention, coherced.

By analogy it is not true that we usually intentionally try to avoid to elicit a bad impression "in order" *not to experience* the unpleasant feeling of shame; we want a good reputation and esteem: this is our motivation. It is false that we avoid to do something bad and unfair "in order" *not to experience* the uncomfortable guilt feeling; we want not to be bad, but to be correct and moral. However, the avoidance of such unpleasant feeling states is there; it possibly is a negative reinforcement of certain actions and, in a sense, our behavior "in order to" avoid them, has such a finality; but it is not—usually—our aim.

3. *Reinforcement Effects on Cognitive Representations*

The other fundamental unification move of behavioristic models and devices and cognitive mental "representations" and processing, is not just to put the two systems in competition or in convergence one with the other, but to say that behavioristic rules also apply to higher level cognitive mental representations and not just to perceptive stimuli and pre-planned executive responses (Bargh and Ferguson 2000; Castelfranchi 2001).

For example, it plays a very crucial cognitive role in the fact that we act on the basis of what we believe, but many of these beliefs are not explicitly formulated or activated, taken into account, and reasoned about. However, they are not challenged ("surprise"), they remain just *presupposed.* There are a lot of "presupposed tacit assumptions" under any action of ours. For example, when I decide to walk in that direction (to go to my office) as usual and routine-like way, not only that I implicitly believed that my office was there (since this was at the beginning—before building a mere routine), but I also "assume" that the floor will support me, that it is safe. I have no reason for thinking about that (consciously or unconsciously), such assumption is not active at all. However, even these presupposed and implicit assumptions (which can also be formulated in a not propositional format i. e. sensory-motor or procedural) if the action succeeds, they get an automatic feedback of confirmation, they are more stable, reinforced ("credible"), and remain presupposed. This also is one of the reasons why failure is a crucial experience for discovering, understanding, and learning.

This *doxastic reinforcement,* the unconscious mechanism is so important in human cognition that it was the advice of Pascal about how to arrive to believe something you cannot rationally believe: you have to act "as if" you believe it, "as if" it was true that… and you will come to believe so. And it is also a classical prescription of cognitive-behavioral psychotherapy in order to abandon some dysfunctional (for Beck[6] "irrational") belief you have: recognize that you can change your mind; *"stop acting or thinking on the basis of the old belief"*, and act in the light of a new belief, and continue to behave in the new way even though it feels phony to act so, and "*that will cause the new belief to become real and a part of your 'natural' behavior"*. This reinforcement effect due to the feedback of a successful action does not only apply to the background (implicit or explicit) beliefs, but also to the adopted plan and means (and to beliefs that are valid), to the goal (by increasing its value as for its attainability and probability). It also reinforces our attachment to our final motivating goals and to our values. Not by reasoned conclusions, evaluations, meta-beliefs but by some sort of "reward" to our assumption, planning, objectives, choices, etc. For example, a successful "action schema" increases—by feedback—its accessibility and affor-

[6] Aaron Beck, the father of "cognitive behavioral therapy."

dance, the probability to be retrieved and chosen next time and some sort of index/measure of its validation. This feedback reinforcement is the fundamental route for their automatization, packing, routinization and habits construction.

A different case is "affect/feeling as information." The normal, canonical cognitivist view is that the cognitive appraisal (beliefs, evaluations) of an event is the forerunner of the emotional response; however, the other way around also exists: feeling something as evidence, as base for believing something. For example, feeling some worry, fear, as a base for *believing* that a threat, some danger is there. Now, given this reverse process the two mechanisms can be combined in a vicious circle (like in panic crisis):

Bel: "There is danger!" ⇒ "Fear" ⇒ feedback *reinforcing* the belief of danger.

4. *Reconciling System 1 and System 2*

First, the conflict between Syst1 and Syst2 is not a matter of a conflict between "rational" or "cultural" aspects against "instinctual" aspects (in case between mere learning by reinforcement vs. true resolutions). Nor is it simply a matter of a conflict between "rational" mechanisms against emotional mechanisms (like in Loewenstein's model).

(i) System 2 is "*reason-based*," that is based on beliefs and evaluations, but this is different from "rational."

(ii) Moreover, both Intentions and activated Classifiers can have *an emotional-impulsive origin*.

Second, the two systems are not just in competition and conflict.[7]

Syst1—it's true—can bypass *deliberation* at all; they compete with each other. Not only "decision" produces action, but also other mechanisms that bypass a real deliberation process:

- reactivity and rule-based behavior
- emotion impulses (like in Lowenstein's view)
- habits and script-based behavior; routines, practices and conformity

But this is not the full story.

Syst1 (with its intuitive, impulsive "values" and "reasons" for preference ("reasons of the heart") and Syst2 (with its reasoned, arguable evaluations and preferences) can interact/interfere with each other, and we can decide by taking into account both: the reasoned values (the reason of the Reason) and the felt values (the heart's reasons) (Castelfranchi 2016). Moreover, Syst1 and Syst2 can be translated one into the other. Many acts originally "driven" by intentions, "in view of," etc. can *become automatic*: routines, habitual, reflex-like, respondent. Classifiers activated by conditions and context, where the original "purpose"

[7] For a deep criticism of the *duality of mind* see also Viale (2019).

remains inactive and implicit. Example: "automatically stop at the red light" that originally when we were learning to drive the car was a real decision. On the other hand, a merely automatic reaction can become problematic, not executable in a given context and we have to reformulate an intention and make a real decision; like at red traffic light but with the siren of an ambulance behind us.

However, the most important form of interaction (not separation) of the two systems and their teleological devices is the fact that *any intentional action (intention to do) when put into execution must be implemented at a lower layer of not really intentional sub-acts*, merely automatically adjusted and just retrieved from our action-repertoire.

The schema we proposed for such integration/implementation is the following one:

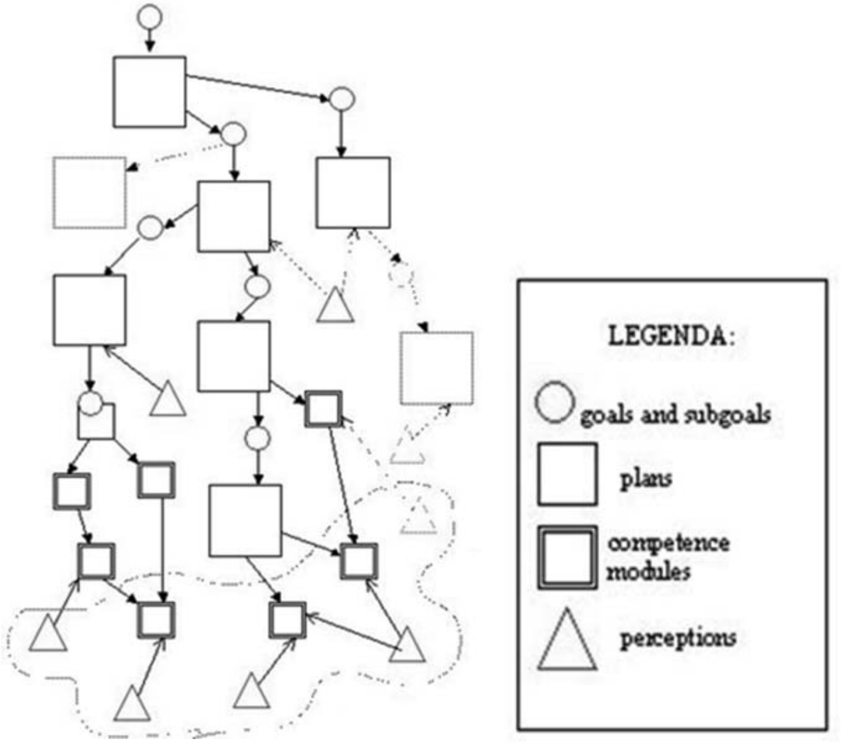

Fig. 3 Functional continuum between Intentional Goals and automatic Classifiers

There is a functional continuum: The top part is more similar to the BDI (Beliefs Desires Intentions) model (Rao and Georgeff 1995; Bratman 1987). The lower part is more similar to Behavior Networks (Maes 1989) and uses anticipatory classifiers (Pezzulo et al. 2008; Pezzulo & Castelfranchi 2009). Executive Intention ("Intentions in action") are/must be *implemented* in lower structures (production rules, reflexes, classifiers), which, when specified, are represented in sensory-motor images/schemes.

For example, the intention to open the door is executed by a lot of micro-actions (bend our fingers, pull, move our feet to pass) which are not "intentional" but finalistic schemas. When I do intentionally take a walk I do not intentionally bend my feet. (See Dunja Jutronić's paper in this volume.)

5. *Considerations on "intentions" in homage to Kathy Wilkes*

"Intentions" only in language using organism? To discuss this thesis we have first to make clear what kind of goal are "Intentions" in our model and where they derive from.

5.1. *What "Intentions" are: a kind of Goal*

"Goals" and "Motives" do not mean "Desires." It is not synonym of "goal" like in Bratman's BDI model (Bratman 1987). Desires are just one *kind* of goal. Desires are endogenous (and usually pleasant) and with "norms" we have just to cut some possible course of action by *making some desire of the subject practically impossible or non-convenient*. Intentions do not derive just from "desires" but also from other kind of goals. They can derive from norms, prescriptions, *duties*; but *"duties" are not "desires";* they are *goals from a different source*, with a different origin: they come from outside (*exogenous*),[8] they are imported, "adopted," they are "prescriptions" and "imperatives" from another agent.

Not all goals have to be "(actively) pursued," like for "intentions." A goal is not a goal only if/when pursued. Some of them (like having a sunny day) are not within our power: to realize them is not up to us, but depends on other "agents" or external forces, thus we cannot really "pursue" them. Other goals are just partially up to us; we have to do something but then the final result depends on the others, or on luck. Thus, we may have actively pursued goals (goals pursued through our active actions), but also merely passive goals; and the latter can be of two very different kinds:

- goals we have just to wait for, to hope for their attainment; which do not depend at all on us: we cannot do anything (else).
- goals whose realization depend on us and on our "doing nothing," that is abstaining from possible interference. We would have the power to block that event/result, and we decide to do nothing in order to let it happen (inaction, "passive action").

Furthermore, because not all goals are directed towards *approaching* a desirable outcome goals can also be directed towards *avoiding* an

[8] However, see later about the internalization of the "authority" and internal moral imperatives.

undesirable outcome (Elliot 2006). Avoidance and approach represent two mental frames, two different psychological dispositions and mind settings (see Higgins' avoidance and approach "regulatory *focus*" in his 1997).

Not all our goals are "felt" because not all of them are represented and defined in a sensory-motor format (see below).[9] The two most important kinds of felt goals are *desires* and *needs*.

Intentions are those goals that *actually drive our voluntary actions or are ready/prepared to drive them*. They are not another "primitive" (like in BDI model), a different mental object with respect to goals. They are just a kind of goal, the final stage of a successful goal-processing with very specific and relevant properties (see Castelfranchi and Paglieri 2007).

In a nutshell, in our model, an *intention* is a goal that:

1) has been activated and processed;
2) has been evaluated as not impossible, and not already realized or self-realizing (achieved by another agent), and thus *up to us*: we have to act in order to achieve it;[10]
3) has been chosen against other possible active and conflicting goals, and we have "decided" to pursue it;
4) is consistent with other intentions of ours; a simple goal can be contradictory, inconsistent with other goals, but, once it is chosen, it becomes an intention and has to be coherent with the other intentions;[11]
5) implies to the agent's belief that she knows (or will/can know) how to achieve it, that she is able to perform the needed actions, and that there are or will be the needed conditions for the intention's realization; at least the agent believes that she will be able and in condition to "try";
6) being "chosen" implies a "commitment" with ourselves, a mortgage on our future decisions; intentions have priority over new possible competing goals, and are more persistent than the latter (Bratman 1987);
7) is "planned"; we allocate/reserve some resources (means, time, etc.) for it; and we have formulated or decided to formulate a plan con-

[9] We mean that, for example, we cannot say "I feel the intention of…" simply because the sensory-motor format of the represented anticipatory state is not specified in the very notion of "intention." "Intention" is a more "abstract" representation, and kind of goal, with a non-specified codification. Looking at a goal as an "intention," we abstract away from its possible sensory components.

[10] An intention is always the intention to "do something" (including inactions). We cannot really have intentions about the actions of other autonomous agents. When we say something like "I have the intention that John goes to Naples" what we actually mean is "I have the intention *to bring it about that* John goes to Naples."

[11] Decision-making serves precisely the function of selecting those goals that are feasible and coherent with each other, and allocating resources and planning one's actual behavior.

sisting of the actions to be performed in order to achieve it. An intention is essentially a two-layered structure:
(a) the "intention that," the *aim,* that is, the original processed goal (for example, to be in Naples tomorrow);
(b) the "intention to do," the sub-goals, the planned executive actions (to take the train, buy the tickets, go to the station, etc.). There is no "intention" without (more or less) specified actions to be performed, and there is no intention without a motivating outcome of such action(s).

8) thus, an intention is the final product of a successful goal-processing that leads to a goal-driven behavior.

After a decision to act, an intention is already there even if the concrete actions are not fully specified or are not yet being executed, because some condition for its execution is not currently present. Intentions can be found in two final and pre-final stages:

(a) *Intention "in action,"* that is, guiding the executive "intentional" action;
(b) *Intention "in agenda"* ("future directed" intentions, those more central in the theories cited of Bratman), that is, already planned and waiting for some lacking condition for their execution: time, money, skills, etc. For example, I may have the intention to go to Capri next Easter (the implementation of my "desire" of spending Easter in Capri), but now is February, and I am not going to Capri or doing anything for that. I have just decided to do so at the right moment; it is already in my "agenda" ("things that I have to do") and binds my resources and future decisions.[12]

5. 2. *"Intentions" only in language using organism?*

A very crucial thesis of Kathy Wilkes is her conclusion that "goal-representation can only be ascribed to language using organism" also due to her caring/stressing the distinction between "intentionality" and "intensionality." This is a crucial distinction. However, I disagree about that conclusion/thesis, which refers to "intentions." My first point is that "mental representations" are not only in linguistic format and based on language. We also have another kind of "mental" representation and mental working, i.e. *mental images and to imagine.*[13]

Also this kind of representations are really semiotic, have their "semantics" (content/object/aboutness). Not only Knowledge (epistemic representations) but also Goals (motivational representations) can be mentally represented *in sensorymotor format, as mental images*. Paradoxically, the example used by Miller, Galanter, and Pribram in their

[12] I would also say that an "intention" is "conscious," we are aware of our intentions and we "deliberate" about them; however, the problem of unconscious goal-driven behavior is open and quite complex (see Bargh et al., 2001).

[13] We know that for Piaget the first level of "intelligence" and thinking is precisely "sensorymotor thinking."

famous book was a nail driven into the wall, where the Goal was a mental Image compared with the perceived one.

However, there may be a possible convergent hypothesis with Kathy Wilkes's challenging claim. As we said, "Intention" in the strict sense belongs to the domain of System 2 and is the result of the "deliberative" processing. It is the result of reasoning, preference and choice *based on "arguments," reasons*, that is beliefs supporting one goal or the other; *it can be explained, discussed.* I can even discuss and argue with my own self; but this doesn't imply that the intended goal/objective itself is formulated in linguistic format.

However, the creation of the "intention" also *entails beliefs about the Agent itself*: my skills, know how: "Am I able to; Do I know what/how to do?". And what about my own mind? Perhaps this self-representation should be expanded: it might entail some *meta-cognitive* representation: not only Beliefs about my mind but meta-Goals.

Since an Intention is a goal about my own agency, my performing an action, it might imply *a goal not only about my doing something but about my having the goal*: a goal about my mind and my commitment. But how can this be formulated, represented a goal about my having a goal, my mind?

5.3. *From "Intention" to "Volition"?*

While the goal of doing/performing a given action can be still formulated in sensory-motor format representation, such goal about my goal/mind reasonably would need a linguistic/communicative representation (a reflexive sociality). This is for me the possible point of convergence/agreement with Kathy Wilkes's thesis.

Not the goal/object of the intention and intentional action is necessary linguistic, it can be merely sensory-motor image (like empting my glass; turning off the stove), but its meta-cognitive, reflexive component is linguistic. *A goal about my mind, my having a goal*, must be represented in an abstract, propositional, conceptual form.

However, I would say that this is no longer just "intention" but it is a "will" and a voluntary action controlled by will; a stronger form of intentionality where I'm *socially influencing my-self* (and language and (self-)mind reading are for that). In fact, the so- called "strength of the will" is my influencing power over my-self: to impose my own self to do something and to be committed, and to control myself.

References

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K. and Trotschel, R. 2001. "The automated will: Unconscious activation and pursuit of behavioral goals." *Journal of Personality and Social Psychology* 81: 1004–27.

Bargh, J. A. and Ferguson, M. J. 2000. "Beyond behaviorism: The automaticity of higher mental processes." *Psychological Bulletin* 126: 925–45.

Bratman, M. 1987. *Intention, Plans, and Practical Reason.* Cambridge: Harvard University Press.

Castelfranchi, C. 1998. "Il nevrotico cripto-utilitarista: contro l'ideologia del 'vantaggio secondario'". *Sistemi intelligenti* 10 (2): 307–314.

Castelfranchi, C. 1999. "La fallacia dello psicologo. Per una teoria degli atti finalistici non intenzionali." *Sistemi Intelligenti* 11 (3): 435–68.

Castelfranchi, C. 2001. "The theory of social functions. Challenges for multi-agent-based social simulation and multi-agent learning." *Journal of Cognitive Systems Research* 2: 5–38.

Castelfranchi, C. 2012a. "Goals, the true center of cognition." In F. Paglieri, L. Tummolini, R. Falcone, M. Miceli (eds.). *The Goals of Cognition.* London: College Publications.

Castelfranchi, C. 2012b. "'My mind'. Reflexive sociality and its cognitive tools." In F. Paglieri (ed.) *Consciousness in Interaction: The role of the natural and social context in shaping consciousness*. Amsterdam: John Benjamins, 125–150.

Castelfranchi, C. 2017. "Goal 'Value': Not just 'Dual' but 'Hybrid.'" In T. Everitt, B. Goertzel, A. Potapov (eds.). *Artificial General Intelligence: 10th International Conference.* Melbourne, 45–54.

Castelfranchi, C. and Paglieri F. 2007. "The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions." *Synthese* 155 (2): 237–263.

Conte, R. and Castelfranchi, C. 1995. *Cognitive and Social Action.* London: UCL Press.

Elliot, A. 2006. "The hierarchical model of approach-avoidance motivation." *Motivation and Emotion* 30 (2): 111–116.

Higgins, E. T. 1997. "Beyond pleasure and pain." *American Psychologist* 52: 1280–1300.

Jutronić, D. 2022. "Intentions and their role in (the explanation) of language change." *Croatian Journal of Philosophy* 22 (3): 327–350

Loewenstein, G. and O'Donoghue, T. 2004. "Animal Spirits: Affective and Deliberative Processes in Economic Behavior." *Microeconomic Theory eJournal.* https:// cpb-us-e1.wpmucdn.com/blogs.cornell.edu/dist/b/5495/files/2015/10/will5_05-227fjlg.pdf

Miller, G.A., Eugene Galanter, E. and K. H. Pribram. 1960. *Plans and the Structure of Behavior.* New York: Henry Holt and co.

Pezzulo, G., Butz, M., Castelfranchi, C. 2008. "The Anticipatory Approach: Definitions and Taxonomies." In G. Pezzulo, M. V. Butz, C. Castelfranchi and R. Falcone (eds.). *The Challenge of Anticipation: A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems.* Cham: Springer 2008, 23–43.

Pezzulo, G. and Castelfranchi, C. 2009. "Thinking as the Control of Imagination: a Conceptual Framework for Goal-Directed Systems." *Psychological Research* 73: 559–577.

Viale R. 2018. "The normative and descriptive weaknesses of behavioral economics-informed nudge: depowered paternalism and unjustifed libertarianism." *Mind and Society* 17: 53–69.

Viale R. 2019. "Architecture of the mind and libertarian paternalism: is the reversibility of system 1 nudges likely to happen?" *Mind and Society* 18: 143–166.