# Against boredom : 17 essays on ignorance, values, creativity, metaphysics, decision-making, truth, preference, art, processes, Ramsey, ethics, rationality, validity, human ills, science, and eternal life to Nils-Eric Sahlin on the occasion of his 60th birthday

Persson, Johannes; Hermerén, Göran; Sjöstrand, Eva

2015

# *Against boredom*

FRI TANKE

# *Against boredom*

*17 essays*

ON IGNORANCE

VALUES

CREATIVITY

METAPHYSICS

DECISION-MAKING

TRUTH

PREFERENCE

ART

PROCESSES

RAMSEY

ETHICS

RATIONALITY

VALIDITY

HUMAN ILLS

SCIENCE

AND ETERNAL LIFE

TO NILS-ERIC SAHLIN

ON THE OCCASION

OF HIS 60TH BIRTHDAY

EDITED BY

JOHANNES PERSSON

GÖRAN HERMERÉN

AND EVA SJÖSTRAND

# Table d'Hôte

# Investigating the development of creativity: The Sahlin hypothesis

INGAR BRINCK

*Abstract*

How should the development of creativity be approached? Many accounts of children's creativity focus on the relation between creativity and pretend play, placing make-believe and the mental exploration of possible scenarios about the world at the fore. Often divergent thinking and story-telling are used to measure creativity with fluency, originality, and flexibility as indicators. I will argue that the strong focus on conceptual processes and higher-order thought leaves procedural forms of creativity in the dark and hinders a proper investigation of the development of creativity. Creativity involves both strategic and procedural elements and the mental and physical manipulation of ideas are equally important. Sahlin's notion of rule-based creativity might serve as the starting-point for an approach to the development of creativity that is neutral as to the underlying nature of creativity and that permits investigating creativity independently of language. On this approach, creativity is characterized by the violation and subsequent replacement of a rule or norm that underlies a given activity with a novel strategy or procedure. When, where, and how children will manifest creativity is conditional on the kind of rule or norm that is violated.

## 1. *Introduction*

Discussions of children's creativity tend to become polarized. Are children naturally creative, or on the contrary, do they need to be educated in creative thinking? Roughly, those who favour the view that children are naturally creative think of creativity as a social and cultural phenomenon that involves imagination and play and starts to develop in the pre-school years (e.g., Garaigordobil & Berrueco, 2011). Those who hold that creativity requires education think of it as a cognitive phenomenon, a property of the individual, that emerges later in childhood and requires training of divergent thinking and logical reasoning (cf. Russ & Fiorelli, 2010). Related but not identical to the second view are the conceptions of creativity as a gift to the happy few and of the creative individual as a genius. These conceptions will not concern us here.

Glăveanu (2011) notices that the first view considers children active and interactive, while the second one pictures them as passive and receptive. A more nuanced understanding of the development of creativity will position itself somewhere in between the two extremes. There is no real contradiction between imagination and cognition; creativity relies on both. Likewise, thinking of creativity as a biological function of the brain similar to memory, attention, inhibition, and anticipation does not rule out that socio-cultural factors influence its functioning or that it may benefit from practice.

## 2. *The relation between creativity and play*

To understand children's creativity it is common to study play. Pretend play in childhood has been shown to affect creativity in adulthood (Russ & Wallace, 2013). Children continually engage in everyday creativity also outside the context of play, e.g., when figuring out a way to train the dog or finding a faster way to get home from school (Russ & Fiorelli, 2010). A major reason why play is considered of central importance to creativity concerns pretend play that involves make-believe and encourages exploring a variety of possible scenarios about the world, such that

build on re-arrangement of known events as well as such that are completely new or novel to the child. Novelty is essential to creativity.[1]

Longitudinal studies reveal that pretend play increases cognitive flexibility in a longer perspective (Russ, 2004). Russ, Robins, & Christiano (1999) found that quality of fantasy and imagination in early pretend play predicted creativity operationalized as divergent thinking over time, independent of IQ.[2] A study by Singer & Lythcott (2004) suggests that when pretend games are encouraged in school as part of the curriculum or during play time this leads to enhanced imaginativeness and, indirectly, creativity.

The experience of free or unstructured play has been demonstrated to have positive effects on originality in subsequent activity, but less on fluency or flexibility as measured by the Torrance Tests of Creative Thinking (Beretta & Privette, 1990). In a study on 6–7 year-old-children, 52 children were split in two groups (Howard-Jones, Taylor, & Sutton, 2002). One group played with salt-dough, the other one did a structured exercise that involved copying text from a board. Then all children were asked to make a collage of a creature with a range of tissue-paper materials. After a few days the experiment was repeated with the groups' changing tasks. Analyses of the children's results by teachers revealed a significant positive effect of preceding tasks upon creativity.

Creativity does not only entail novelty, originality, flexibility, and divergent thinking (cf. Brinck, 1997; Sahlin, 2001). The research on creativity in adults stress that

---

1. Boden (1991) makes a useful distinction between person-related and historical novelty. The former kind concerns novelty in relation to the person (persons) who has generated the idea. Then the idea is known by otherpeople and does not appear creative from their perspective. Everybody can be creative in the person-related sense. The latter kind concerns novelty in a larger context, where the outcome is truly novel and of historical importance. It requires expert knowledge in the field to which the idea pertains.

2. Divergent thinking is the elaboration of ideas in many different directions. It is used in brain-storming, a technique or method for the free generation of alternative ideas.

creative ideas also are productive: Once generated, ideas are evaluated according to how likely they are to result in a proper solution or answer, one that actually will work. Evaluation involves refinement of the idea along different lines (Brinck, 2003). There is no reason to demand less from children. Creative ideas that emerge during play often are produced under pressure to maintain play in the face of unexpected difficulties, and must be adequate to do their job. The problem has to be addressed on the fly or play comes to an end. One example is when the children who are playing are of different ages and therefore have different understanding of what is going on, another when too many repetitions within the same group of children has made the theme of the play (say, to play doctor) predictable and boring, fostering negative emotions and attitudes. In a group of children that play together often, conventions (Lewis, 1969) emerge for how to deal with such interruptions. In contrast, a situation that is new to the children and they don't know how to deal with, calls for creativity.

Mottweiler & Taylor (2014) notice that although elaborated role play (pretending in which children imagine and act out the part of another individual on a regular basis) is considered an early indicator of creativity, there is a lack of evidence of a relation between it and performance on creativity tasks during the preschool years. They maintain that the measures of creativity that are commonly used such as divergent thinking tasks are not appropriate for young children, because generating multiple solutions to the same problem is unfamiliar and cognitively challenging for them. This remark points to the importance of developing tests that have ecological validity. Accordingly, Mottweiler & Taylor developed two new measures of creativity based on a storytelling task, in which 4- and 5-year-old children were asked to complete a story, and a drawing task, in which the children were asked to draw an imaginary person. They showed that the children who engaged in elaborated role play had higher creativity scores on both measures (controlling for age and language ability).

Glăveanu (2011) argues that children develop creativity in interaction with adults and through play and experi-

mentation with cultural artefacts. He highlights that creativity develops over time, and that how it is expressed depends equally on the socio-cultural environment and the particular scaffolding of the individual child.[3] This means that children who grow up in same socio-culture in the end may display different forms of creativity and to different degrees. The education and pedagogy they receive most likely will differ between individuals, as will the socio-economic status (SES) of their families (SES is measured as a combination of education, income, and occupation). These factors tend to influence children's possibilities to engage in free play, e.g., their motivation and preparedness as well as the amount of time they are allowed for it. However, we cannot draw the conclusion that children from families with low SES will not be creative. There may be other ways to develop creativity than in free, imaginative play, ways that reward originality and novelty in the concrete, so to speak. In the next two sections, I will present a broader conception of creativity than found in much of the research on children's creativity.

### *3. Creativity is procedural and strategic*

Mottweiler's & Taylor's (2014) object to the use of the divergent thinking paradigm for testing creativity in pre-schoolers. Yet it is not certain that measures of creativity that rely on story-telling or narrative will do better. The younger the children, the less reliable their narratives will be as indicators of creativity, because young children have not yet acquired sufficient linguistic proficiency for expressing their creativity verbally in a consistent and reliable way. Furthermore, not all forms of creativity depend on language, which means that measures that rely on verbal reports may overlook subjects who are creative

---

3. The term "scaffolding" means there is a single more knowledgeable person, usually a parent, who helps the child to develop new skills by giving the support the child needs to perform a certain task or reach a goal (Wood, Bruner, & Ross, 1976). Once the child has learnt how to perform the behaviour, the scaffolding is removed.

but whose linguistic skills are less than average (e.g., for socio-economic reasons, or because they have an impairment that affects language use). Finally, certain forms of creativity may be difficult to express and analyse verbally. Skill-based creativity that relies on knowhow and bodily experiences is not readily accessible by verbal means (Brinck, 2007). Brinck (1999) refers to such forms of creativity as procedural and describes them as embodied, situated, and interactive.

Procedural creativity makes use of contextual information for taking cognitive short-cuts. Strategic creativity is conceptual and context-independent, and therefore can release the subject from states that hinder free association and fluency, e.g., functional fixedness. Brinck (1999) maintains that creativity contains both procedural and strategic elements. In this respect, creativity seems similar to expertise. Höffding (2014) observes that the skilled coping of experts such as chess players, musicians, and athletes is phenomenologically complex and spans both absorption and reflection. Höffding bases his argument in an extended case-study of the expertise possessed by the members of a string quartet.

A large part of the creative process takes place in the external world and consists in thinking with external models (Brinck, 2003, 2007; Fioratou & Cowley, 2009). Evaluative judgments are prompted directly by perceptual information and visuo-spatial reasoning (Weller, Villejoubert, & Vallée-Tourangeau, 2011). The information that moves the creative processes in one direction as opposed to another may not reach conscious awareness. Except for perception and sensory-motor information, affect plays a central role for procedural creativity. Rietvald (2008) explains the unreflective skilful action of expert craftsmen in terms of the notion of situated normativity. He argues that a particular type of affective behaviour is essential for evaluation without reflection (for "getting things right"), described as a reaction of appreciation in action. To conclude, while it would be wrong to contest the value of narrative as a tool for investigating creativity, in certain circumstances a measure of creativity that does not rely on language may be more appropriate.

Conceiving of creativity exclusively along the lines of make-believe or pretence and the capacity for exploring a variety of possible scenarios about the world suggests that it is essentially conceptual or representational and involves a more or less deliberate or conscious 'juggling' with alternative realities. Such a conception of creativity has been related to capacities for theory of mind and thinking about other people's 'inner worlds'.

It is hard to deny that imaginative play that involves social role-taking depends on understanding that people can take different perspectives and that their thoughts and experiences may differ (Singer & Singer, 2005). This does not prove that creativity depends on theory of mind. Perhaps both creativity and play depend on some other more general function that supports flexibility. Moreover, it is not clear that all forms of pretence involve role play. Pretence does not always concern living (or phantasy) creatures. Equally, it is uncertain that creativity as a rule comprises perspective-taking in the sense in which the research on theory of mind defines perspective-taking.

Physical play, e.g., ball play, hide and seek, and building castles in the sand, huts in the wood, or towers and cities with Lego or other kinds of physical objects, also depends on imagination and on envisaging alternative, sometimes quite complicated scenarios. Physical exploration and the trying out of possible or alternative actions in contexts of instrumental action contain the playful manipulation of ideas – not conceptually, but as embodied in or exemplified by artefacts. Because the result of physical manipulation reveals itself directly to the senses and feedback is immediate, the actions of idea generation, exploration, testing, and evaluation tend to co-occur or overlap. Certain problems are better dealt with in physical space than conceptually in imagination, and the testing and evaluation of ideas then can be over in a few seconds. Software designers, architects, craftsmen, developers (and players!) of computer games, and fashion designers are just a few examples of professionals who organize the creative process around the manipulation of objects (and ideas) in space and time, physically or virtually, and let it be guided by sensorimotor processes rather than conceptually (Brinck,

2007; Gedenryd, 1998; Kirsh & Maglio, 1994; Wynn, 1993).

There is a test that acknowledges that creativity can be processed and expressed by bodily actions and movement: Torrance's Thinking Creatively in Action and Movement (TCAM). It uses movement and manipulation exercises to test creativity in children age 3 to 8 years and comprises four activities. Three of these consist in generating alternative ways of performing an action. The test is designed to measure fluency, originality, and imagination. Because the subjects are not asked to express their creativity verbally, the test has the advantage of being independent of the verbal skills of the subjects. However, like many other tests of creativity, TCAM conceives of creativity as a form of divergent thinking that involves perspective-taking and perspective change. It is questionable that creativity boils down to the capacity for seeing things from different perspectives. The central thing is to see or do things in a novel way – not in an alternative way.

*4. Approximate problem-solving*

Sahlin (2001) gives numerous real life examples of creativity that together demonstrate the complex character of creativity and that creativity occurs in quite diverse situations. I will present four instructive examples. The first example concerns Admiral George Rodney. He defeated the French in the battle of Les Saintes 1782 by deliberately neglecting certain of the British army's Fighting Instructions that regulated how to perform a battle at sea. This unexpected strategy was inspired by a book on naval tactics based on the author's experiments as a boy with toy boats in the garden pond.

Second, the artist Dan Wolgers had been booked to have an exhibition at Gallery Lars Bohman in Stockholm. He came up with the idea of delegating the task of producing the exhibition to an advertising bureau instead of doing it himself. He showed up at the opening to see his work for the first time. Wolgers' behaviour caused a big scandal that reached far beyond the usual art crowd. In breaking the rules for how to prepare an exhibition and

what it means to exhibit art, Wolgers raised fundamental questions that rarely are addressed about what art is and what an artist is and should do. For instance, in what ways can the assistants (that many contemporary artists have) help the artist in creating his works of art and how much can they do while remaining assistants?

Third, Richard Fosbury won the Olympics in high jump in 1968 using a new way of jumping that came to be known as the Fosbury flop. Instead of running towards the bar, jumping with his front facing it, Fosbury turned his back towards the bar before jumping. It took him 5 years to develop his style to perfection and win the Olympic Gold medal. Already 4 years later at the next Olympics a number of athletes copied his way of jumping. Nowadays almost everybody jumps with the back towards the bar. The Fosbury flop originated partly by chance. Fosbury had difficulties with the prevalent technique. He felt that he needed to raise his hips not to knock down the bar. When he did so, he automatically started to drop his shoulders and lay back. The resulting flop was as a consequence of how the human body is built.

The final example concerns Theresa Berkley who ran a flagellatory brothel in England in the beginning of the 19th century. She is famous for the invention of the "Berkley Horse", a triangular frame to which a person can be tied in any desirable angle for flogging. It was a great success. Sahlin (2001) describes Berkley's capacity to change her expectations about what flogging means and break with the values of her time as typical of creative people.

Sahlin's examples illustrate that creativity is deliberate and purposive and that it requires quite extensive knowledge or skills in the field it concerns. The chance that a mere guess will be creative is next to zero. More importantly, they draw our attention from divergent thinking and imagination to problem-solving. In all four cases, there is a problem to be dealt with, or, what amounts to the same thing, a question to be answered: How can the French be defeated? How can I make an exhibition that is not conditioned by contemporary theories and norms about art? What other ways are there to improve my results in high-jumping than quantitatively (by increasing

my training)? How can I improve the competitiveness of my business by meeting the demands of the buyers?

I suggest that we conceive of creativity as problem-solving – in a broad sense. As opposed to regular problem-solving that is exact and fixed, creative problem-solving is approximate. That it is approximate means that it is unclear how the problem can be solved and what the solution might be. Conceiving of creativity as approximate problem-solving minimizes the risk for making premature or arbitrary assumptions about its nature, e.g., by defining it in terms of behaviour that presupposes certain types of cognition and hence by definition confines creativity to agents that have the required cognitive capacities. This gives the present suggestion an advantage over views that conceive of creativity in terms of divergent thinking or imagination.

Whenever a question needs answering, an issue needs to be sorted, a goal needs to be reached, a task needs to be performed, or an idea needs to find an expression, and the subject does not know how to do or even what to do, then the situation calls for creativity – whether in the domains of science, art, culture, sport, or of any everyday activity such as cleaning, cooking, gardening, or shopping (Brinck, 1997). In principle, any issue can be a problem in the broad sense (as you may have experienced in daily life) – how to graft fruit in the absence of the right material, how to build a hut for your kids in the woods without the proper tools, how to account for the origin of life, or how to get to a meeting in time in a foreign city when facing a wild strike in the public transportation system.

Creativity is an open-ended process that is useful when a method or procedure for solving the problem is unavailable. It is unclear what your options are. You don't know how to proceed or go about and, moreover, cannot anticipate the result of your inquiries. Consequently, creative problem solving is not algorithmic or guaranteed to lead to a solution, but makes use of 'informed guesses' and heuristics or rules of thumb that often are implicit.

## *5. Rule-based creativity: The Sahlin hypothesis*

I have argued that creativity is not limited to certain domains, activities, or behaviour, and that it comprises both conceptual and sensorimotor processes. This must be taken into account when investigating its development. But if creativity is pervasive and comes in a wide variety of guises, what unites it? What does it consist in? Sahlin (1991) provides a simple and ingenious answer to these questions. He distinguishes between two fundamental types of creativity. Concept-based creativity consists in inventing new concepts that change our perception and understanding of a phenomenon. Rule-based creativity consists in breaking the rules that underlie an activity and inventing new strategies or procedures for how to approach it.

In the rest of the article I will briefly outline how Sahlin's notion of rule-based creativity may be spelt out to serve as the basis for empirical investigations of the development of creativity in children and adolescents, alongside other techniques that tap into verbal and conceptual skills such as narrative, divergent thinking, and free association. One important advantage of Sahlin's definition of creativity is that it emphasizes a central characteristic of creativity: novelty. The ability to generate a great number of alternative ideas (and see things from different perspectives) is of less significance to creativity than the ability to invent novel ways of perceiving or acting. It is enough to produce one novel idea. Number does not count.

The rule-based approach to the development of creativity takes for granted that children are sensitive to norms and rules and the ways that norms and rules simultaneously circumscribe and enable behaviour in daily life. These assumptions are uncontroversial, but need to be made more specific to permit working out how the notion of rule-based creativity can be used in empirical work. For instance, we need to determine what it means to be sensitive to a rule or norm and what the behavioural criteria are.

It is possible to discern a few trends in the research on children's understanding of rules and norms in develop-

mental psychology. For instance, it has been shown that before the age of 4 years, children have difficulties following abstract rules and more easily get distracted by features that are irrelevant for performing the task. They can know a rule but be unable to apply it (Towse et al., 2000; Zelzo, Frye, & Rapus, 1996). The executive function and capacities for perspective-taking of preschool children are not yet fully developed, which hampers abstract reasoning and cognitive flexibility. Other studies show that 3–6-year olds can endorse a norm of fairness verbally but neglect it in practise, because although they understand its appropriateness, they are not personally motivated by it (Smith, Blake, & Harris, 2013). Furthermore, there is evidence that younger children use rules for predicting others' behaviour but by 8 years, like adolescents and adults, they tend to base their predictions on the individual preferences of others (Kalish & Shiverick, 2004). Finally, it has been shown that 3-year-olds understand the nature of constitutive rules, which define and support arbitrary social activities (games of chess and monopoly, or sports like ice-hockey and tennis) as well as social institutions and functions (the government, church, school, police, queen, etc.) (Rakoczy, 2006; Rakoczy, Warneken, & Tomasello, 2008).

Children operate with a number of more or less distinct concepts of rule and norm. The data suggest that there is not one single developmental path for the understanding of rules and norms, but several different paths that each roughly corresponds to a particular type of rule or norm. As a consequence, granted that creativity consists in the breaking or violation of rules and norms, it can be expected to emerge at distinct points in development depending on what kind of rule or norm is violated. That is, on the Sahlin hypothesis, although rule-based creativity consists in the same type of behaviour across contexts and ages, performance and quality is conditioned by whether it involves the violation of, e.g., moral or social norms, conventions, rules of logic, or constitutive rules. Children develop an understanding of rules and norms piecemeal, certain types being mastered at an earlier age than others. Thus it seems that this view would allow for precise pre-

dictions of when in development creativity will emerge relative to the particular type of norm or rule the transgression concerns. To exemplify, creativity in domains that require using abstract rules or logical reasoning to solve a problem might be expected to occur in middle or late childhood.

The present approach has several advantages. First, the focus lies on novelty as opposed to variation of ideas, hence on quality, not quantity. Second, in testing whether the subject actually can provide a new strategy or procedure for solving the problem, it puts the weight on the result of the creative process. This stands in contrast to approaches that test whether the subject would be able to provide alternatives, i.e., whether the subject has the capacity for generating many ideas or, say, for divergent thinking. That a subject has imagination does not imply or guarantee that she can come up with an idea that works. This means that the present approach examines whether subjects in fact are creative as opposed to examining whether they have the capacity for being so. Third, rule-based creativity can be conceptual or representational as well as experiential or sensorimotor, and so explains creativity globally, whatever the domain (theoretical physics, engineering, chess, sports, craft, cooking, et cetera). Forth, the rule-based approach acknowledges that both bodily and psychological processes can generate creative ideas and so agrees with recent data that suggest that sensorimotor and cognitive processes interact in the creative process. Five, the rule-based approach can be used to explain creativity in subjects of any age and in any context.

### *6. Identifying creativity: behavioural criteria*

Empirical investigation of creativity presupposes that there are objective criteria that make it possible to decide whether certain behaviour is creative or not. To establish such criteria, we first need to clarify what it means to break a rule (violate a norm) in the present context. Obviously, mere neglect or disregard of a rule is not creative. The point is to break the rule for a purpose, i.e., to replace it

with behaviour that may contribute to solve the problem. Doing so involves recognizing that the existing rule is wrong and that it needs replacement by another behaviour that is at least more likely than not to solve the problem. This raises the further question whether the new behaviour must be successful to be creative.

The definition of rule-based creativity does not mention that the novel behaviour must be successful to be creative. Yet, it is possible that at a certain age children consistently will display the required behaviour, viz., they invent new strategies for approaching the activity, but nevertheless they fail to solve the problem. They then would be expected to produce strategies that lead to positive results only later in development. This would mean that the behaviour is complex and that it comprises something more than the mere ability to break rules with the goal of improving one's strategy or heuristics. I suggest that this 'something more' concerns the ability to replace the rejected rule with an *efficient* action or set of actions. Most probably, doing so would comprise evaluating the action(s) relative to the estimated end state while working it (them) out, something that seems to require at least roughly anticipating the consequences of the action(s). Such a procedure would sort out inefficient actions, but it cannot guarantee that the remaining action(s) actually will be successful.

We have reached the point where we can formulate four behavioural criteria that permits identifying a subject as creative according to Sahlin's definition of rule-based creativity:

(1) the subject does not engage in the expected behaviour A
(2) the subject produces another behaviour B
(3) the subject has not engaged in or encountered behaviour B before (at least not in similar circumstances)
(4) behaviour B can lead to (or: leads to) a solution to the problem

Behaviour A = a rule or norm

The third and fourth criteria each have a weaker and a stronger reading and further analysis would be needed to settle which readings are correct. Subjects that satisfy all four criteria are creative. In contrast, a subject that satisfies the first, second, and third criterion has limited understanding of the behaviour that underlies creativity, and does not know how to produce a strategy or procedure that is both novel *and* successful. An alternative interpretation is that (given that she satisfies the first three criteria) the subject might in fact be able to solve the problem, and therefore is creative, but her behaviour is not reliable (over time). She cannot be relied upon to provide strategies or procedures that lead to a solution (but she may do so once in a while). I will leave it to the reader to decide which interpretation (if any) is preferable and why.

## *Funding*

This work was supported by The European Science Foundation grant number 429-2010-7181.

## *References*

Beretta, S. & Privette, G. (1990), Infuence of play on creative thinking. *Perceptual and Motor Skills*, 71, 659–666.

Boden, M. (1991), *The Creative Mind*. New York: Basic Books.

Brinck, I. (1997), The gist of creativity. In A. E. Andersson & N.-E. Sahlin (Eds.), *The Complexity of Creativity*. Synthese Library Vol. 258, Dordrecht: Kluwer Academic Publishers, pp. 5–16.

Brinck, I. (1999), Procedures and strategies: Context-dependence in creativity, *Philosophica*, 64(2), 33–47.

Brinck, I. (2003), Evaluation and testing in creativity. In Rojszczak, A., Cachro, J.,and Kurczewski, G. (Eds.), *Philosophical Dimensions of Logic and Science*. Synthese Library, Vol. 320, Dordrecht: Kluwer Academic Publishers, pp 331–344.

Brinck, I. (2007), Situated cognition, dynamic systems, and art. *JanusHead*, 9(2), 407–431.

Fioratou, E. & Cowley, S. (2009), Insightful thinking: Cognitive dynamics and material artifacts. *Pragmatics and Cognition*, 17(3), 549–572.

Garaigordobil, M., & Berrueco, L. (2011). Effects of a play program on creative thinking of preschool children. *Spanish Journal of Psychology*, 14(2), 608–618.

Gedenryd, H. (1998), *How Designers Work*. Lund, Lund University Cognitive Studies 75.

Glăveanu, V.P. (2011, *Children and creativity: a most (un)likely pair? Thinking Skills and Creativity*, 6(2), 122–131.

Howard-Jones, P.A., Taylor, J.R. & Sutton, L. (2002), The effects of play on the creativity of young children, *Early Child Development and Care*, 172(4), 323–328.

Høffding, S. (2014), What is Skilled Coping?: Experts on Expertise. *Journal of Consciousness Studies*, 21(9-10), 49-73.

Kalish, C.W., Shiverick, S.M. (2004), Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, 19, 401–416.

Kirsh, D., & Maglio, P. (1994), On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18, 513–549

Kloo, D., Perner, J., & Giritzer, T. (2010), Object-set-shifting in preschoolers: Relations to theory of mind. In B. W. Sokol, U. Müller, J. I. M. Carpendale, A. R. Young and G. Iarocci (Eds.), *Self- and Social-Regulation: Exploring the Relations between Social Interaction, Social Cognition, and the Development of Executive Functions*, pp. 193–217. Oxford: Oxford University Press.

Lewis, D. (1969), *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Mottweiler, C. M. & Taylor, M. (2014), Elaborated role play and creativity in preschool age children. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 277–286.

Rakoczy, H. (2006), Pretend play and the development of collective intentionality. *Cognitive Systems Research*, 7, 113–127.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008), The sources of normativity: Young childrens' awareness of the normative structure of games. *Developmental Psychology*, 44, 875–881.

Rietvald, E. (2008), Situated Normativity: The Normative Aspect of Embodied Cognition in Unreflective Action. *Mind*, 117(468), 973–1001.

Russ, S.W. (2004), *Play in child development and psychotherapy*. Mahwah, NJ: Earlbaum.

Russ, S. & Fiorelli, J. (2010), Developmental Approaches to Creativity. In J . Kaufman & R. Sternberg (Eds.) *The Cambridge Handbook of Creativity*, 233–249. New York: Cambridge University Press.

Russ, S.W., Robins, A.L., & Christiano, B. A. (1999), Pretend Play: Longitudinal Prediction of Creativity and Affect in Fantasy in Children. *Creativity Research Journal* 12:129–39.

Russ, S.W., & Wallace, C.E. (2013), Pretend play and creative processes. *The American Journal of Play*, 6(1), 136–148.

Sahlin, N.-E. (2001), *Kreativitetens filosofi*. Nora: Nya Doxa.

Singer, D.G. & Singer, J.L. (2005), *Imagination and play in the electronic age*. Cambridge, MA: Harvard University Press.

Singer, J.L. & Lythcott, M.A. (2004), Fostering school achievement and creativity through sociodramatic play in the classroom. In E.F. Zigler, D.G. Singer & S.J. Bishop-Joseph (Eds.) *Children's play: The roots of reading*, pp. 77–93. Washington DC: Zero to Three Press.

Smith, C.E., Blake, P.R., & Harris, P.L. (2013), I Should but I Won't: Why Young Children Endorse Norms of Fair Sharing but Do Not Follow Them. *PLoS ONE*, 8(3): e59510.

Towse, J. N., Redbond, J., Houston-Price, C., & Cook, S. (2000), Understanding the dimensional change card sort: Perspectives from task success and failure. *Cognitive Development*, 15(3), 347–365.

Weller, A., Villejoubert, G., & Vallée-Tourangeau, F. (2011), Interactive insight problem solving. *Thinking and Reasoning*, 17(4), 424–439.

Wood, D., Bruner, J. S., & Ross, G. (1976), The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 17*(2), 89–100.

Wynn, T. (1993), Layers of Thinking in Tool Behavior. In K. R. Gibson and T. Ingold (eds.), *Tools, Language, and Cognition in Human Evolution*, Cambridge: Cambridge University Press.

Zelazo, P. D., Frye, D. & Rapus, T. (1996), An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.

# Known unknowns and proto-second-personal address in photographic art

LINUS BROSTRÖM

For reasons that are more coincidental than interesting, if those were the only two things they could be, I have recently given some thought to the work of a couple of (so called) art photographers and, in particular, why this work is so compelling. Numerous artists would actually fit the bill when it comes to the issue I am going to address, but I have in mind particularly the work of the very well known Robert Adams, as well as the work of the lesser known (of the two) Raymond Meeks. Adams, of course, is the highly influential documenter of the often reckless transformation of the American West, offering at the same time a broader reflection upon man's relation to the environment and, ultimately, a reflection upon what to do with what one is given. Raymond Meeks, on the other hand, has typically found his subjects in the more intimate surroundings of his and his family's own home, even though there are enough obvious affinities with Adams' work to place him in the same broad tradition – what can be heard here are different dialects rather than different languages. Actually, Adams too has made arrays into the more intimate, and some of the work of his that speak to me the most shares with Meeks' this interest in small things close to home, registered in a way that is both tentative and tender. I am going to take as a given in what follows that this is high quality work, worthy of our undivided attention. Conceivably someone might take issue

with that assumption, but that beef would have to be taken outside, and preferably at some other point in time. So why are these images so strong? Or more cautiously, if you insist, what would account for their strength, if indeed they did possess it? The truer such an account rings, the better it might establish its explanandum – or so one might hope, in a leap of (let's call it) coherentist abduction.

Let us take a quick look at two small but, I think, telling examples from the photographic oeuvre of Adams and Meeks. Adams' *Questions for an overcast day* was published in 2007 by the Fraenkel Gallery. It is a rather small volume, with 33 plates picturing fragments of young Oregon alder trees, and nothing else; their twigs and lacerated leaves refusing to display any discernible order in the wind. The only text in the book consists of these questions:

> What would account for the condition of the leaves – droughts, insects, rocky ground, disease, herbicide, wind?
>
> Are the leaves beautiful?
>
> Is there something wry in the hieroglyphics? And something humorous about a person taking photographs, the camera hand-held, between gusts of wind?

The images are characterized by their compositional restraint, with little or no attempt by Adams to simplify for effect or a comfortable viewing. He does nothing to downplay the complexity of the motif, and does not in any other way bring to these alders the photographer's standard tools aimed at purifying what is in the frame, for a clean enough statement. As the book progresses we get closer and closer to the leaves, structures being revealed that were hardly visible from the more distant perspective, and vice versa. Aside from this zooming in (or zooming out, if read from back to front), the differences between the images may seem small. No new ideas appear to be introduced; it is simply more of the same, one might say. But these "snapshots" are emphatically not containers of well-defined ideas in the first place; on the contrary, they are at heart exploratory, and the photographer has deferred to the images themselves considerable responsibility for what they (one by one or collectively) convey.

Many of Raymond Meeks' limited edition artist books embody a similar spirit. For example, in *Pretty girls wander*, from 2011, Meeks may make use of more overt symbolism and traditional photographic storytelling than *Questions for an overcast day* does, but just like Adams, he is so clearly asking questions rather than making (emphatic) claims – and not rhetorical questions, but ones that seem deeply felt, and find their final shape on the pages of the book, not before. The images tell a story of a child having grown up and being ready to move away from the family to a larger world, not breaking ties but becoming less susceptible to the gravitational forces of the first home. We get to feel the bodily presence of that daughter who is being drawn towards independence, the familiar houses moving out of the frame, the promise of the railway tracks, and the withering flowers as a possible punctuation mark. There are also two silver halide prints included the book, where the transparency of the physical medium contributes in a beautiful way to the theme. Highly metaphorical, yes, but crucially, just what these metaphors say about home, belonging and dependence Raymond Meeks does not know; the images, through their *tone*, if nothing else, makes it clear that the artist is still unresolved about these larger questions.

Now, just what the relation is between photographic art, on the one hand, and truth claims or epistemic claims, on the other hand, would be a hugely challenging topic, and one for other occasions. It would be surprising if there was no connection whatsoever between the two. At the end of the day a purely non-cognitivist approach to much photographic art may prevail, but even such an approach would have to account for common everyday intuitions about the representational claims of those branches of photography that on some level at least seem to traffic in documentation. It is advisable not to make too much of photographic and other ostensibly representative art's similarity to enterprises that clearly are in the business of aiming for truth. Still, whether such art is somehow hypothesis generating, hypothesis testing, or related in some other way to the search for truth are surely among the questions that we should address when trying to understand how pictorial art works.

Photographic art may make claims, may even express knowledge, but also on that assumption the artist would do well to know his or her epistemic limitations. Obviously art may turn out unsuccessful in any number of ways, but one nearly foolproof way of failing, I contend, is for the artist to know just a bit too much beforehand, and being eager to convey whatever it is that he or she claims to know. When the photographer simply tries to channel through the photographic medium whatever he or she has already identified as a certain truth, the artist typically exhibits two different vices: plain arrogance, and a failure to trust the medium. Arrogance, because who could feel that certain about the difficult subjects that art finds worthy of putting on the table? Inappropriate distrust, because surely the artistic expression itself should be expected to do much of the work. While there are uncountable examples of convincing so called conceptual art, there are perhaps even more examples of conceptual art manifesting this particular sin. At its worst, such art is not only boringly didactic, but articulates whatever rough ideas it has on its mind in ways that could have been much better phrased otherwise – it simply becomes bad philosophy. In contrast, while Adams and Meeks certainly have some idea about what matters, they feel their way to the destination. They would not dream of exploiting those tried alders or fragile backyard memorabilia for the mere promulgation of some Very Important Message. Whatever lessons may be there, they are not known to the photographers beforehand, if at all.

This tentativeness is not to be conflated with being lackluster or without a vision. Humility and precaution can both be traced back to the same virtue of paying attention to epistemic risks, but they are not the same thing. Adams, Meeks and others like them are undeniably expressing a point of view, suggesting what conditions we face as human beings, and what is worth caring about. Photography that makes no suggestions about anything would have little to say for itself. There is indeed art, including photography, that throws all precaution to the wind, making much more controversial statements than the work of Adams and Meeks does. Some high quality art goes out on a limb and

makes very bold conjectures while other high quality art is more provisional, making smaller gestures. The present point is not that the latter is superior to the former. Compelling art may be as Popperian as it likes in presenting a clearly falsifiable perspective. The point is rather that when bold refutable artworks are presented, these are unlikely to be aesthetically convincing unless the artist recognizes them as bold and refutable, as attempts to capture that which is often elusive.

Now, in research and real-world decision making there is much at stake, and it is imperative to manage epistemic risks in a prudent way. Here inattention to what is still not known, or to what might after all turn out false, may ultimately cost us plenty. Whether there are in art the same kinds of instrumental reason for handling epistemic risks in appropriate ways is contentious, but whatever one's view on the stakes, epistemic hubris typically makes for *uninteresting* art; knowing what's unknown will at least be an aesthetic virtue. The degree to which the photographic work speaks to us, as recipients, often covaries with the artist's refusal to confidently transmit pre-formulated messages, or so I suspect. This may well generalize to other artistic mediums, incidentally. Thomas Hampson, barytone extraordinaire and one of the wiser pedagogues in classical song, makes an analogous point, for example, when he insists that in order for a vocal performance to be convincing, the singer should never ever try to *project* whatever meaning is allegedly found in a libretto to the audience, but try to get to know the character, hear the phrase before it is sung, "breathe into" that, and "make it audible". Who knows, in opera too, perhaps, it ultimately comes down to the virtue of acknowledging what one just may be wrong about.

I have sided with those who believe that being too confident about what to convey is detrimental to photographic art, whereas a certain kind of modesty makes for more convincing expressions. On second thought, epistemic humility on the part of the artist may in fact be secondary to a different distinguishing characteristic with which it quite reliably covaries. Rumor has it that correlation ≠ causation, and possibly one should look elsewhere

for that good-making characteristic that makes the photography of Adams, Meeks and others working in the same spirit special. This other characteristic has to do with the artist's *address*.

While photographers and other artists would do well not to dwell too much on how their art will be received, they typically hope for some reaction. Now, most everyone sees the difference between addressing somebody as a You and attempting to influence an individual in other ways. Stephen Darwall has rather recently built a comprehensive normative theory around the differences between these kinds of address (see e.g. his 2006 *The Second-Personal Standpoint*, on Harvard UP). The example used by Darwall to introduce the relevant distinction concerns a, broadly speaking, moral issue, viz. what reasons there might be to stop stepping on another person's foot. The first kind of reason consists of your recognizing that the person might be in pain (or suffer in some other way), that you are in a position change that, and that there are moral norms governing all individuals in your position to the effect that, all else being equal, inflicting pain upon somebody should be avoided. This is what Darwall would call a third-personal reason (for action). A different reason for you to remove your foot would be if the person whose foot you are standing on simply asked you to. The *second*-personal standpoint, Darwall stipulates, is the standpoint we take when in this fashion we make direct claims upon one another's will, regardless of what independent reasons there might be to meet the demands or expectations expressed.

Art, when successful, is typically not in the business of making moral prescriptions, but unless it somehow connects with moral concerns, albeit in the most indirect ways, it is hard to see why we should bother. Some art seemingly attempts to contribute third-personally, as it were. It makes prejudices visible, questions deeply entrenched norms, scrutinizes distributions of power, etc. by redirecting our attention and doing its best to take advantage of our synaptic plasticity. Although one might be uncomfortable with the negative connotations of the word, there is a clear sense in which this kind of art operates through *manipula-*

*tion*; by going for effects on our sensibilities or states of mind. Photography has its fair share of those kinds of artists, even though there wouldn't be much point in listing names. Other artists contribute not so much by pushing and pulling levers but by talking to us, as equals, and sometimes in a hoped for conversation. (Some art combines in interesting ways aspects of the two approaches; Marina Abramovic's "The Artist is Present" would perhaps be a case in point.) Adams, Meeks and others who resemble these two in their approach to the artistic act do seem to me to work their magic in part by having us feel addressed as a Me, rather than someone whose outlook could be altered through artistic engineering. In presenting to us those messy grey alder leaves, or that family life that will slowly dissolve into vague memory, Adams and Meeks seem to hand over the relevant visions for our competent consideration, rather than to manufacture experiences. Now, speaking to someone as a You, in the relevant sense, is not think of and direct oneself to any one individual in the artistic process, much less deliver a communiqué directly in the hands of that addressee. It's not about projecting, as we said. Rather, it is to hold oneself accountable to anyone listening (viewing the images) in roughly the way one would when meeting in person.

Well, is it really? Isn't this notion of photographers addressing unknown consumers of their art as a You stretching things beyond what is intelligible? It probably is. So let me modify the suggestion: What characterizes the modus operandi of Adams, Meeks and their likes is perhaps not so much true second-personal address, but that they *lament* the fact their address, through their images, cannot be genuinely second-personal. Infusing all this work is at least a *wish*, of sorts – naïve as it may be – to connect with the addressee as equals do in authentic second-personal dialogue.

I no longer read much poetry. But as for many other young adults who have had even the slightest inclination to look in that direction, poetry was once part of my literary diet, and unsurprisingly Tomas Tranströmer was one of those who made an impression. I am now reminded of an older poem of his, "To friends behind a frontier" ("Till

vänner bakom en gräns"), here in Robin Fulton's translation:

1
I wrote so meagerly to you. But what I couldn't write
swelled and swelled like an old-fashioned airship
and drifted away at last through the night sky.

2
The letter is now at the censor's. He lights his lamp.
In the glare my words fly up like monkeys on a grille,
rattle it, stop, and bare their teeth.

3
Read between the lines. We'll meet in 200 years
when the microphones in the hotel walls are forgotten
and can at last sleep, become trilobites.

There is an obvious political comment being made here, about oppressive surveillance societies, but personally I read this poem just as much as an expression of frustration over the author´s, by necessity perhaps, inability to speak to the reader as friends do. As a matter of fact, I believe the greatness of a great Tranströmer poem depends as much on his *address* as that breathtaking imagery that he is so famous for. But never mind. The images of Robert Adams and Raymond Meeks could be trusted companions of yours, if you let them. I raise these questions about the role of uncertainty and proto-second-personal address in their photography as questions for an overcast day.

# Critical moral thinking without moral theory

JOHAN BRÄNNMARK

We need critical moral thinking. It might not be a need that we think about much in everyday life, and even if we see the need we might not deeply feel the need. Nevertheless we need it simply because moral values and ideas play such a huge role in our lives, greater than we might perhaps think since much of the work they do lies in the background. We do not, for example, have to think about whether or not to commit murder, an action physically possible for most of us most of the time, we simply do not even consider it as an option. The options we do consider have already been filtered and screened in a way that leaves most actions not even on the table. Morality is of course not the only filtering and screening mechanism, but it is a central one.

Moral theory is an area where we find critical moral thinking, but does our need for critical moral thinking translate to a need for moral theory? It will be argued here that not only do we not need it, there are positive reasons for eschewing it. The point is not that there is no use for moral philosophy, but rather that moral philosophy would be better off without moral theory, at least as it is standardly understood. Before getting to the reasons for this, there is however a need to briefly go through that standard understanding.

## 1. *Moral theory and the question of method*

The target of the argument here is this thing called "moral theory", but if it is to be possible to hit that target we need a reasonably clear picture of it. Since moral theories in some form or other have been formulated for at least two centuries, and perhaps even two millennia, one should not expect to be able to capture something like the essence of moral theory in only a few brief sentences. For present purposes it will have to be enough to identify some very common ideas and then target these. Here are four such very common ideas which can be understood as providing us with desiderata for a successful moral theory:

(1) It seeks the underlying justifications for all of our more concrete and particular judgments about right and wrong (at least the reasonable ones).
(2) It seeks to systematize these underlying justifications, or ultimate normative concerns, into a coherent set of principles.
(3) It seeks to increase our understanding of morality by allowing us to identify a central core of morality and allow us to see how more concrete applications are related to that core.
(4) It seeks to provide practical guidance by allowing us to critically reflect on the principles and precepts on which we act and ultimately replace our starting set of principles and precepts with a superior set.

While some moral theories are ultimately boiled down to a criterion of rightness, i.e., a set of necessary and jointly sufficient conditions for an action being right, not all moral theories reach that level of systematicity, although arguably it can still be seen as a desideratum in the pursuit of traditional moral theory. There might in the end be practical reasons against actually employing such a criterion in everyday deliberation, for instance because the calculations involved in applying the criterion to particular cases would be far too complex and cognitively demanding. It can instead be used to reconstruct the rules of thumb we

do use in our deliberations. This is a common stance among consequentialists with respect to their favored criterion of rightness.

Moral theory is a branch of critical moral thinking and as such it tends to start out with our everyday moral ideas (so-called common-sense morality), and then work its ways towards a better state of ideas. Even philosophical intuitionists like Henry Sidgwick and W. D. Ross take common-sense morality quite seriously. This is not surprising: if human beings have the capacity to discern moral truths through some form of intuition it would perhaps be somewhat surprising if everyday morality did not have some kernel of truth to it. Not all moral theorists are intuitionists, however, and among coherentists there is an even greater reliance on the moral views that we already have. The standard form of something like coherentism in ethics is the reflective-equilibrium procedure that was developed by John Rawls in his 1971 book *A Theory of Justice*. Ultimately the idea of reflective equilibrium should perhaps be seen as methodological rather than epistemological, but the core idea is at any rate that we are to start our process of revisionary moral thought by identifying two different sets of ideas. On the one hand we have our considered moral judgments, i.e., moral judgments that we have a high degree of confidence in, and on the other we have different principles. These two sets are unlikely to fully cohere and what we can then do is work our way towards a state in which we might have rejected some considered judgments, but where we have also reached a set of principles that strongly coheres with the remaining considered judgments. There might of course be several candidates as to which set of principles that gives us the highest degree of coherence (at the least cost), but this problem is hopefully to be settled through debate between moral theorists of different persuasions.

The method of reflective equilibrium can be specified in different ways, but one thing is central to it: the role of principles. The main challenge against moral theory in the philosophical literature of the last thirty-five years or so has come in the form of moral particularism and in the associated idea of a holism of reasons, where the former is

the rejection of the strong role for principles assumed by most moral theorists while the latter provides a rationale for this rejection, namely that features that count as normative reasons are not invariant, i.e., they might count for an action in one context, against in another, and not at all in a third. If holism about reasons is correct then generalizing the moral concerns we have into meaningful principles will be at best difficult, but perhaps not even possible. The earliest formulation of this approach came from John McDowell, but the leading proponent is probably Jonathan Dancy. It should however be said that to some extent particularists, at least those that are holists about reasons, still share a certain picture of moral deliberation with moral generalists/theorists. On that picture, our judgments about right and wrong should ultimately be traceable to a set of reasons for and against the actions considered. These reasons are usually thought to be in a form that is structurally apt for generalization, e.g., "that this action involves killing is a reason against performing it". The moral generalist will think that if we have correctly identified the relevant moral concerns, they can then be generalized into principles, e.g., "whenever actions involve killing there is a reason against performing it", whereas a moral holist will think that no such meaningful generalizations can be made. There are many things to say (and to be honest many of them have already been said in the existing literature) about the dispute between the particularist and the generalist, but here the focus will be on another question: what if already the picture that they share is a faulty one?

## 2. *The nature of moral deliberation*

Moral deliberation clearly involves processes of categorization in terms of placing persons, states, events, and actions into normative categories such as "right" and "wrong" or "good" and "bad", all closely related to our decisions about what to do from a moral point of view. The element of decision and action involved in moral choices certainly distinguishes moral thinking from many other processes of thought also involving categorizations,

but this does not remove the fact that we are still dealing with mechanisms of thought that revolve around categorizing things we encounter in different ways.

As anyone familiar with the literature on categorization will know, there is no consensus on how categorization works, even though (or perhaps because) there has been a great amount of research in this area, especially in the last 35 years or so. The most sensible position given this state might simply be that there is no single way in which categorizations function; but even given such an ecumenical stance, there will always remain a question with respect to different areas of thought which mechanisms operate there, to what extent any of them dominates, and whether they play distinguishable functional roles. So even though one might not ultimately have to choose between them, it might still be helpful to start by outlining some of the main approaches in the field. Here are four of them:

(1) *The Classical Theory*: processes of categorization are about determining whether an item satisfies the list of individually necessary and jointly sufficient conditions for being a member of the category in question.
(2) *The Prototype Theory*: processes of categorization are about determining whether an item to a sufficient degree has properties that are typically possessed by members of the category in question.
(3) *The Exemplar Theory*: processes of categorization are about determining whether an item is similar enough to specific known members of the category in question.
(4) *The Theory Theory*: processes of categorization are a form of inferences to the best explanation, where placing the item belonging to a certain category provides an understanding of the item in the context where it is encountered.

There are a couple of things that should be noted here. One is that the classical theory is very much a minority approach within psychology, linguistics, and cognitive

science today, at least as a general theory about categorization. Another is that these theories are theories in a very loose sense, they might be (and have been) developed in different directions. But while there is much to say about the similarities and differences between them, the important question for present purposes is how they might bear on philosophical work on clarifying or improving our mechanisms of moral categorization. One important line then runs between (1) and (4) on the one side and (2) and (3) on the other. Given the former two there is an inherent striving towards systematization and coherence built into the very possession of a category, whereas on the latter two, being in possession of a category might be something mainly associative and fairly inarticulate. Adherents of traditional moral theory will presumably want to claim that moral categorization is in line with (1) and (4) because in that case our mechanisms of moral categorization, the ones that are to be fine-tuned through a reflective-equilibrium process, will be structurally similar to what the end product of our theorizing is supposed to look like. On the other hand, if these mechanisms are in line with (2) and (3) instead, this would spell trouble for that process since the starting materials would then be structurally very dissimilar to what the end product should look like. Categorization is in that case not a matter of falling under a rule, but of having a short enough distance to an exemplar or prototype.

On the basis of the above we might distinguish between two ideal types of moral deliberation. On the one hand we have the *Rulebook Model*, according to which moral deliberation revolves around rules and principles or specific identified features of a situation that serves as reasons for or against preforming different actions (and where these features are in a generalizable form although they might not ultimately be generalizable into exceptionless rules). On the other hand we have the *Storybook Model*, where moral deliberation is governed by exemplars and/or prototypes in the form of a network of stories, persons, and fairly concrete examples of actions where whether an action that we contemplate is categorized as right or wrong depends on where it is located in this conceptual

space in terms of its distances to relevant prototypes/exemplars. Information in the Storybook need not be encoded in abstract terms but can be in the form of images and metaphors as well. Relations between different objects in the Storybook will to a great extent be associative rather than logical, which is also what will make them considerably less fit for generalization and theory construction.

On the Rulebook Model, deliberation would largely be a two-stage process where we first identify actions that are open to us and then deliberate in terms of the reasons for and against choosing these different actions. In the end, one action emerges as the one favored by the balance of reasons. On the Storybook Model, we will first locate the situation that we face in the conceptual space of prototype/exemplar situations that constitute our moral understanding, and depending on how it is categorized, different actions will emerge as the one to perform, usually without any idea about why beyond a sense that they *fit* the situation. Given this kind of model, then if there is a clear prototype/exemplar action that emerges out of this process we might not actually consider any reasons (for or against) at all before we decide how to act. One's sense of having discovered how to act will be more like having found the missing piece of a puzzle than having deduced a consequence from a more general principle. Which of these two models is most truthful to actual moral deliberation? In the end we probably exemplify both of them, but which one is more important? If the answer is that the Rulebook Model holds the most truth, then we can reasonably see Storybook elements of our deliberations as something that through translation can be integrated into theories; if the Storybook is dominant, then this process of reconstruction is starting to look like a much more difficult enterprise and our ambitions for theory construction should in all likelihood be seriously downsized.

Now, if we look at human psychology more broadly, it is commonly thought that higher-order human thought can be divided into two main types of processes. First we have what is often called System 1 thinking, which is fast, automatic, and non-conscious. The links made between different items tend to be associative and the items them-

selves are often metaphors, images, and narratives. Reactions tend to be rooted in affect and gut feelings, often related to previous experiences. Then we have System 2 thinking: slow, deliberate, and conscious. The connections made between items here tend to be logical and the items themselves are abstract symbols, words, and numbers. There is conscious assessment of evidence for believing different things and options tend to be evaluated in terms of articulated ideas about values and norms. Within this broader picture it seems clear that the Rulebook belongs to System 2 and the Storybook to System 1. If moral deliberation is similar to other forms of thinking, this would mean that the Storybook is dominant simply because System 1 is the dominant form of thinking in everyday life due to its much higher speed and more modest requirements of informational input.

If we look at moral philosophy in general then clearly the Rulebook Model is the favored form of thinking so the above conclusion is one that moral philosophers are likely to resist, but if we look honestly at the situation, setting vested interests to the side, it seems difficult to avoid the conclusion that we should expect the Rulebook to play only a marginal role in our actual moral deliberations. And while phenomenology is always problematic to appeal to, at least this writer finds that this is a conclusion that fits with his own phenomenological understanding, and that one must be very much in the grip of some theory if one finds that weighing reasons is something that dominates our moral deliberations. If we turn to moral psychology, we will also find that this kind of picture is corroborated by what is perhaps the most influential work of the last 10–15 years, namely that of Jonathan Haidt, who has argued that when reasons do enter in it is mainly in the form of post-hoc rationalizations that we engage in when being questioned about our choices, i.e., when we are actually deciding what to do we tend to employ System 1 thinking whereas when we are questioned we engage in System 2 thinking. It is not that there are no Rulebook elements in our moral thinking, but they are the exception, not the norm.

### *3. Implications*

Where does the above conclusion leave us in terms of how critical moral thinking should be conducted? One possible response is simply to still insist on the traditional form of moral theory while acknowledging that the task might be very difficult, but that nevertheless it is the path we must take. Yet while there might certainly be a role for principles to play, if nothing else because they are highly communicable, a more realistic appraisal of the situation might be that the distance between where we are at the starting-point of everyday moral deliberation compared with the supposed end-point of highly abstract and general moral theory á la Kantianism and utilitarianism is simply too great to be bridgeable. We can certainly formulate such theories, but after two centuries of that kind of theorizing it does seems as if every single specific formulation of such theories is out of fit with at least some aspects of everyday moral thinking. So maybe it is time to give up on the idea that all the intuitive responses of everyday thinking can ultimately be captured in the One Big Theory.

If we look at the history of moral philosophy, then clearly there are philosophers who have made important contributions to moral thought without formulating anything like standard moral theories, Aristotle and Hume being two examples that come to mind, so putting an end to moral theory does not mean putting an end to moral philosophy. Since principles and core values will continue to play a role in public deliberations there can still be contributions to be made both in clarifying them and in shifting emphases between them. This is still a viable undertaking even if one eschews the systematicity of moral theory. At the same time, we should not expect simply to return to something very similar to what Hume and Aristotle did. Given that the Storybook Model captures a great deal of our actual moral deliberations, then one could not seriously engage in critical moral thinking without engaging with the narrative elements of our processes of moral thought. Ultimately, how we as actual human beings understand and apply ideas like justice, autonomy, well-being, and so on, will never just depend on abstract

definitions of these notions but also on the examples that we associate with them. This means that if we want to shift our patterns of application in relation to these notions we must also engage with examples not just as illustrations or arguments but as guiding ideas, where moral reform can very well come through shifts in the examples that we rely on in our moral thinking, sometimes perhaps even without the associated general principles changing in how they are worded or emphasized. For example, abstract criteria of what counts as autonomous choices might be important, but in the end what actual people will classify as autonomous choices will probably to a large extent revolve around the prime examples of such choices that they associate with the notion. As philosophers we cannot change this, because it is simply a matter of how people (and philosophers are people too) function, so if we want to effect moral change we must also work on shifting people's associations, not just providing more sharply defined criteria. Unfortunately, at least at the moment, the examples that philosophers put their best efforts into tend to be more or less outlandish outliers designed to test our theories rather than good examples designed to support good practices.

# Vad är ett missförhållande?

MARTIN EDMAN

Försöket att eliminera ett missförhållande ger ibland upphov till nya missförhållanden. Behandlingen av sjuka patienter kan leda till att de blir ännu sjukare. På skolområdet har reformerna avlöst varandra i en allt snabbare takt. Varför? Vad är ett missförhållande? Frågan är lättare att ställa än att besvara. Klart är att ett missförhållande är något oönskat som bör undanröjas. Dess dålighet är inte en subjektiv uppfattning hos den som drabbas utan något *verkligt dåligt*. Missförhållanden uppfinns inte utan upptäcks. Det som uppfattas som ett missförhållande behöver inte vara det och det som inte uppfattas som ett kan vara det. Under 1700-talet ansågs kaffedrickning vara mycket skadlig för hälsan och det utbredda bruket av kaffe ett missförhållande. På 40- och 50-talet ansåg man att DDT var ett ofarligt och effektivt medel mot insekter. På 60-talet framkom att det i själva verket var fråga om ett allvarligt missförhållande.

Allt som är verkligt dåligt är inte ett missförhållande. Att vi en gång skall dö brukar de flesta anse vara något verkligt dåligt utan att för den skull hålla för ett missförhållande. Missväxt, översvämningar och naturkatastrofer är det inte heller. Anledningen är uppenbar. Ett missförhållande är något verkligt dåligt som på grund av sin dålighet inte endast bör utan även kan åtgärdas. Kikhosta och giktattacker är missförhållanden eftersom de kan förebyggas genom vaccination eller intag av allopurinol. Trafikolyckor orsakade av att man är onykter eller inte använder säkerhetsbälte är också missförhållanden. Att vi alla skall dö är inte ett missförhållande endast så länge det inte är känt hur man gör för att få evigt liv.

Missförhållanden har *en homogen etiologi* som gör dem möjliga att eliminera med en enda åtgärd. Olyckor som orsakas av att man inte använder säkerhetsbälte eller är onykter kräver två skilda åtgärder och är därför två missförhållanden. Kikhosta och gikt är inte en utan två sjukdomar.

Eftersom missförhållandena är den delmängd av de verkliga dåligheterna som kan åtgärdas, finns två sätt att upptäcka ett missförhållande. Det kan, som i DDT-fallet, ske genom att en given företeelses verkliga dålighet avslöjas eller genom att en åtgärd som undanröjer företeelsen blir tillgänglig.

### *1. Subsidiära effekter*

Men allt som är dåligt, har en homogen etiologi och kan åtgärdas är inte ett missförhållande. Den som köper aktier riskerar att kursen faller. En elitidrottsman som tränar hårt riskerar att skada sig. Det är i båda fallen något dåligt och oönskat som kan åtgärdas genom att man avstår från aktieköp respektive intensiv träning. Men kursförluster och idrottsskador är trots detta inte missförhållanden. Den som placerar i aktier har som överordnat mål att öka sitt kapital. Den som tränar hårt har som överordnat mål att bli bäst. I båda fallen är man medveten om att det uppsatta målet är förenat med risker. Aktieköparen vet att det är ofrånkomligt att kursen på aktier ibland faller. Idrottaren vet att det är ofrånkomligt att hård träning ibland ger träningsvärk och skador. Sådana ofrånkomliga sidoeffekter är inte missförhållanden utan *subsidiära effekter* till ett överordnat gott.

Det finns alltså två slags dålighet, verklig dålighet och *tolerabel* dålighet. Den tolerabla dåligheten är något som man måste stå ut med för att undvika den verkliga dåligheten, nämligen att inte uppnå det överordnat goda. Huruvida en verklig dålighet är ett missförhållande eller ej bestäms av om det finns eller inte finns en känd åtgärd. Huruvida en dålighet är verklig eller tolerabel är ett faktum och påverkas inte av vad som är känt och okänt.

## *2. Uppkomsten av missförhållanden*

Att åtgärda en tolerabel dålighet i tron att den är en verklig dålighet är något verkligt dåligt. Om kunskap är ett överordnat gott i skolan och det krävs läxor och betyg för att nå kunskap och dessa tas bort, nås inte det överordnade kunskapsmålet. Åtgärden har skapat ett missförhållande. Om ordnad migration är ett överordnat gott som medför en risk för uppslitande avvisningar och dessa inte verkställs går det överordnat goda förlorat och ersätts av oordnad migration. Om en oundviklig sidoeffekt till ett överordnat gott blockeras, blockeras även det överordnat goda och ett missförhållande uppstår.

Vad som i en given situation är överordnat gott och vad som är subsidiära effekter är fakta och blir kända genom att undersökas. Om det hävdas att något som är mycket vanligt och funnits sedan länge dels är mycket oönskat och dels enkelt kan åtgärdas finns det anledning att ta ett steg tillbaka och fråga sig om det oönskade kan vara en subsidiär effekt till ett överordnat gott. En sådan undersökning avslöjar att katederundervisning och läxläsning är tolerabla dåligheter för skolbarn. Med skicklighet och lite tur hade den visat att insektsplågan är en tolerabel dålighet i förhållande till DDT i naturen.

## *3. En filosofisk fråga*

Existerar det verkligen något sådant som ”verkligt dåligt” och ”överordnat gott”? Enligt en vanlig uppfattning bestäms vad som är bra eller dåligt av den sociala kontexten och ändras över tid. Det som är heligt i en kultur är en hädelse i en annan. Om svaret på frågan är nej förlorar diskussionen om missförhållanden fotfästet.

Det finns två huvudtyper av svar. Den ena är att företeelser inte är goda och dåliga i sig. En saks värde bestämmer inte individens val utan det är individens val som bestämmer sakens värde. Under 1700-talet ansågs exempelvis det utbredda bruket av kaffe vara ett ytterst allvarligt folkhälsoproblem av samma slag som rökning idag anses vara. En hälsoivrare skulle bestämt protestera mot detta och hävda att rökning verkligen är skadlig och att

den rökning som, trots varningstexter och information om rökningens skadlighet, pågår inte är ett inbillat utan ett verkligt missförhållande. När rökarna drabbas av allvarliga hälsoproblem ångrar de nämligen sitt rökande och slutar. Problemet är att det inte gäller alla. Precis som det finns bergsklättrare som trots att de råkat ut för en fallolycka *väljer* att fortsätta att klättra finns det rökare som *väljer* att bortse från det allvarliga hälsoproblemet och fortsätter att röka.

Den andra huvudtypen av svar på frågan är att det är värdena som bestämmer valen och inte tvärtom. Det som skiljer den oförbätterlige klättraren och inbitne rökaren från oss andra är inte deras val av värden utan valet av väg till ett gemensamt, överordnat och i sig gott mål. För att det är så talar att deras handlande uppfattas som ett olöst problem. De inbitna rökarna och klättrarna antas befinna sig på en för oss okänd väg mot ett överordnat mål, relativt vilket olyckorna och hälsoproblemen är subsidiära sidoeffekter. När vi identifierat detta mål och denna väg förstår vi varför de handlar som de gör. Vi skulle ha gjort samma vägval om vi befunnit oss i deras situation. Det är med andra ord värdet som styr valet och inte tvärtom. En meningsfull diskussion om överordnad godhet och verklig dålighet är möjlig. Eventuella oenigheter om vad som är verkligt dåligt och överordnat gott är skenbara. Om diskussionen misslyckas beror det på deltagarnas bristande fantasi och fattningsgåva och inte på deras fria vilja.

### *4. Läkemedel och missförhållanden*

Neurosedyn utvecklades på 50-talet av läkemedelsföretaget Grunental och marknadsfördes som ett sömnmedel. Företaget drog, efter att ett stort antal missbildade barn fötts, motvilligt in medlet 1961. Detta ledde till att kraven på godkända läkemedel skärptes. För att godkännas måste ett läkemedel numera genomgå en omfattande och kostsam testprocedur med flera administrativa kontrollstationer. Verksamheten drivs av kapitalstark läkemedelsindustri som producerar läkemedel som behandlar sjukdomar där efterfrågan täcker utvecklingskostnaden och ger vinst. Detta är bra men kanske inte tillräckligt bra. Utvecklingen

av nya läkemedel mot sjukdomar som är ovanliga, framtida eller drabbar fattiga i utvecklingsländerna har bromsats upp. Är detta ett missförhållande eller en oönskad men ofrånkomlig subsidiär sidoeffekt av ett överordnat och gott mål?

Det överordnade och goda målet kan inte vara att skydda alla patienter från biverkningar och skador. Patienter som lider av sjukdomar som är ovanliga etc. skyddas visserligen på detta sätt från biverkningar och skada, men trösklar som bromsar utvecklingen av nya läkemedel mot dessa sjukdomar är det sista de önskar sig. Risken att nya läkemedel mot deras sjukdomar har biverkningar är uppenbarligen en subsidiär sidoeffekt till det överordnat goda att de överhuvud existerar och kan utvecklas.

Det överordnade målet kan inte heller vara att förhindra att läkemedel introduceras och marknadsförs på ett lättsinnigt och ansvarslöst sätt. För de som lider av en ovanlig sjukdom är frånvaron av lättsinnig och ansvarslös marknadsföring en klen tröst om detta beror på att det inte finns några läkemedel att marknadsföra.

Är det överordnade och i sig goda målet att skydda den grupp av patienter som lever i i-länder och lider av vanliga sjukdomar mot läkemedelsbiverkningar och skador på bekostnad av läkemedelsförsörjningen för den grupp som lider av sjukdomar som är ovanliga, framtida eller drabbar fattiga i utvecklingsländer? Det är svårt att se hur det skulle kunna vara det.

Det verkar inte finnas något överordnat och gott mål som motiverar de uppställda kraven. Dessa förefaller vara ad hoc åtgärder riktade mot något oönskat och felidentifierat. De vidtagna åtgärderna har träffat en tolerabel dålighet och skapat ett missförhållande.

All behandling med läkemedel är förknippad med en risk för biverkningar och skador. När man behandlar en sjukdom balanseras kravet att inte skada mot kravet att bota och lindra. Balansen ser ut på ett sätt för ett nytt läkemedel mot sömnproblem och på ett annat för ett nytt läkemedel mot en allvarlig och hittills obotlig sjukdom som Creutzfeldts-Jakobs. Biverkningar som för en sjukdom är ett missförhållande är för en annan en subsidiär effekt och en tolerabel dålighet. Vad som i varje enskilt fall

är ett överordnat gott och en tolerabel dålighet beror på den behandlade sjukdomens karaktär, vilka behandlingsalternativ som finns och de allmänna omständigheter under vilka behandlingen äger rum.

Ett exempel kan åskådliggöra problematiken. Det är känt att björksocker in vitro dödar cancerceller och inte nämnvärt skadar normala celler. Det finns omfattande kunskap om hur substanser som socker verkar i kroppen, hur de omsätts metaboliskt och om vilka biverkningar och skador som kan förekomma. Situationen är med andra ord bäddad för att med sedvanlig forskaretik och metodik börja undersöka om björksocker faktiskt kan användas som ett läkemedel vid cancersjukdom. De rådande kraven på godkända läkemedel gör dock detta i praktiken omöjligt. Ett godkänt läkemedel måste kunna patenteras för att täcka kostnaden för godkännandet. Det är svårt att patentera en naturligt förekommande substans som björksocker. Regleringarna fungerar här, precis som för sjukdomar som är ovanliga, framtida eller drabbar fattiga, som ett hinder för att nå ett överordnat och gott mål och är således ett missförhållande.

### *5. Avslutning*

Missförhållanden är kopplade till hierarkiska strukturer uppbyggda av påverkbara delsystem. Den mänskliga kroppen med dess organstrukturer och sjukdomar, ett trafiksystem med dess olika delar och trafikstockningar, hela samhällen med deras organisationer och intressekonflikter är exempel på sådana strukturer. Varje delsystem har som överordnad godhet sin egen funktion och som verklig dålighet sin icke-funktion. När alla delsystemen fungerar, fungerar helheten. Ett missförhållande uppstår när något av delsystemen påverkas så att det inte fungerar. Det inträffar när den fungerande helheten innehåller något som vid första påseende förefaller oönskat men som i verkligheten är en subsidiär effekt till en överordnad godhet. Om en sådan tolerabel dålighet uppfattas som ett missförhållande och åtgärdas får det återverkningar hela vägen upp i hierarkin. Man får vårdhem fyllda av patienter som, medan deras hålbrickor fylls på med medikamenter, blir

allt sjukare och sjukare. Man får missriktade ad hoc regler som bromsar och snedvrider utvecklingen av nya läkemedel. Bakom åtskilliga av dagens missförhållanden skymtar ett *urmissförhållande*: en allmän övertygelse om att det är möjligt att identifiera och framgångsrikt åtgärda något oönskat utan att först förstå hur den helhet är beskaffad som det oönskade är en del av.

# Rambling on the value of truth

PASCAL ENGEL

A cat is objectively valuable

*Ayn Rand*

Although it has become a bit old-fashioned to use this kind of language, it is natural to think that Logic, Aesthetics and Ethics are "normative sciences", and to consider that they deal respectively with the values of Truth, Beauty and Goodness. Ramsey, however, for one, was not convinced that the correspondence is exact in the case of logic:

> For whereas the chief question in Ethics is undoubtedly "What is good?", and in Aesthetics "What is beautiful?", the question "What is true?" is one which all the sciences answer, each in its own domain, and in no way the particular concern of Logic. What Logic studies is not so much the truth of the opinions, as the reasonableness of arguments or inferences. (Ramsey 1991, 3)

Ramsey then hints, in the introduction to his unpublished manuscript *Truth*, that questions of value are to be answered through a psychological investigations about the kind of attitudes which are the source of these values, and in the case of logic about the nature of our opinion and judgments as psychological states and about their rationality in inferences. He was, in other words, a non-cognitivist, and, given his famous view that "there is no separate question of truth, but only a question about the nature of judgment" (Ramsey 1990), a non-factualist both about truth and about the value of truth.

I have learnt most of what I know about Ramsey from Nils-Eric Sahlin. Although for long I have been sympathetic to Ramsey's view on truth and for his non-cognitivist stance on values, including epistemic, I have now come to doubt that they are correct. I try here to give some of the reasons why I prefer a cognitivist conception of the value of truth.

## *1. Truth as prima facie valuable*

If we want to ask in what sense truth is valuable, we should attend some familiar distinctions about values (see e.g. Mulligan 2009). We ascribe to certain objects certain value properties (good, bad, beautiful). But to what kind of entities? What are the bearers of value? Objects? State of affairs? The more the values have content, the "thicker" they are, by opposition to "thin" values. Something can be a value or a disvalue in itself, or in relation to something else. We can conceive of truth as valuable in itself, as a final value, or in relation to another value, as an instrumental value. A value can be intrinsic, when the value is to be found in an object or property in itself, or extrinsic when the value is relative to another object. Something has an intrinsic value when it is valuable for its own sake, and an instrumental value when it is valuable for the sake of something else. There are also various kinds of values: practical, moral, epistemic, aesthetical, social, possibly others. What kind of properties are value properties? Do they form an exclusive kind or do they reduce to another more fundamental kind? In other words do value form a special domain, the domain of the axiological? Or do they have strong connexions, and possibly are reducible to other normative properties such as the deontological ones or in the sense of being things for which we have reasons? When we ascribe value properties, do these properties denote a certain kind of entity – values – or are these properties a projection of our psychological attitudes – of our valuings? Ontologically speaking one can be a realist or an anti-realist about value. Finally one can take value properties to be reducible to natural properties, or to supervene upon these, or not. None of these various issues are

independent from each other. I cannot hope to deal with all these distinctions, but we can try to apply these to the familiar idea that truth is valuable, hence at least a value property.

Although the fact that truth is valuable is a property of our ordinary concept of truth, it is not easy to specify what this property is, in what sense it is a value property and what its bearers are. It is often said that truth is a value. But of what is it a value? Truth by itself, as a property of our beliefs or assertions, has no value and is neither good nor bad, neither beautiful nor hideous. That grass is green or that manganese has atomic number 25 are truths is a fact about these sentences or propositions, and there is nothing valuable in that they are true or describe what is the case. Facts or truths as such do not have any value. If these propositions can be valuable or can have a value, it is as potential objects of our beliefs or of our assertions. Truth is a value property of our beliefs and assertions, which are its primary bearers. Moore said in *Principia Ethica*: "I cannot at any given moment distinguish what is true from what I believe" (Moore 1903, § 80). Commenting this passage, Marian David (2012) proposes the following test. I present you with a list of propositions and ask you: "Mark the ones that are true!" You comply. Imagine now that, concerning the very same list of propositions, I had asked you: "Mark the ones that you believe!" You would have marked the very same propositions. It would seem that if the possession of truth is valuable, the views which associate intrinsically the nature of truth to its possession by a believer will say that truth as a property is valuable. Thus verificationist views, which say that truth is warranted assertibility, coherentist views, which take truth to be coherent belief, or pragmatist theories, which take truth to be a property of successful beliefs, will readily associate the value of truth to some epistemic property. But we should be cautious here to: that truth is valuable relative what we say or what we believe does not entail that truth is an epistemic property. There is no reason to presuppose a form of anti-realism or epistemicism about truth when we attribute to true beliefs a value. The intuitive association noted by Moore between belief and truth

does not even begin to indicate that being true entails or is equivalent to being believed. On the contrary it would seem that in order to be able to ascribe a value to truth, truth has to be a property which is independent of our believing anything about it. In particular the most radical of all epistemic theories, relativism, entails that truth cannot be a value. For in order to accept the idea of the value of truth, or of its disvalue, false belief must be possible. But relativism, or at least the crudest version of this doctrine, does not make room for false belief: according to it all our beliefs are equally true, just in virtue of being *our* beliefs. So all of our beliefs, if simple relativism holds, ought to be valuable. But if all beliefs are equally valuable, how can truth be a value? It cannot accommodate the idea of a value of truth in any objective sense.

If truth is a value property of our beliefs, it is presumably a "thin" and not a "thick" property, as many philosophers since Aristotle have claimed. Even Aristotle's famous "definition" of truth in *Metaphysics* 1025b – "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true" – which is often interpreted as a first statement of the correspondence theory of truth, does not say very much. It is actually a platitude, which features among those which are said to be associated to our ordinary *concept* of truth (Wright (1992): *transparency* ("'P' is true" says the same thing as "P"), *embedding* ("that P is true" can be embedded in other contexts), *correspondence* ("'P' is true if P corresponds to the facts, to reality, to how things are), *objectivity* (truth *contrasts with justification*, is *stable* and *absolute*). To these platitudes one can add that truth is, as William James puts it, "the good in the way of belief": it *is good*, or *better* to have true, rather than false beliefs. It seems that without all these features, including the last one, our concept of belief would not be the one it is.

From the fact that truth is, on the face of it, a thin concept, does it follow that it is a thin value property of our beliefs? Not necessarily. Actually if being valuable is one of the "platitudes" which are attached to the concept of truth, it is not clear that this concept is so "lightweight" (Engel *to appear*). Certainly we do not seem to say very

much when we say that truth is valuable because it is the goal of inquiry or what we "aim at" when we believe, and that error is what we try to avoid. We can express this positive and this negative goal respectively as:

(TG) (i) To believe P if P is true
(TG) (ii) Not to believe P if P is false, and not to believe not P if P is true

But, as a large literature shows (see e.g. Bykvist and Hattiangadi 2007, David 2012, Chan 2013), these goals are not easy to interpret.

First, although it seems obvious that truth is what we try to get when we believe, it is not obvious that our aim should be to believe everything that is true. There are so many truths which are trivial or uninteresting, or dangerous to believe, that we ought to at least qualify (i) by saying that truth has only *prima facie* a value for our beliefs. Second, if it is a value is it a final value, or one which is only instrumental to something else? This question cannot be separated from the following: what kind of value is truth? This seems to depend upon the kind of goals one has when one tries to reach truth. If one is a scientist – if one attends primarily epistemic value – presumably truth is a kind of final, or intrinsic value. But for many practical purposes – if one attends practical values – truth seems to be of merely instrumental value. Things, however, are more complex, for in a number of circumstances, there are conflicts between theoretical and practical values. Can the former trump the latter and vice versa? Fourth, should we interpret (i) and (ii) as specifications of a value property of truth, or as specifications of other kinds of normative properties? Some (Wedgwood 2002, Boghossian 2003, Engel 2004) take truth to be a norm of belief, or a standard of correction of our believing in some constitutive sense. Is the truth of a belief something that we value, or something that we ought to attend to or to conform to? Is it something which we have most *reason* to attend or to conform to? It's one thing for truth to be what it is correct to believe and to be what we aim at. The very nature of the normative concepts that we use here makes a lot of

difference if we want to specify the nature of the relation of truth to belief. Is truth a value *at all*?

## 2. *The eudaimonic value of truth*

If truth is value (*qua* value of true belief) it can be either an intrinsic or an extrinsic value, and it can either be a final or an instrumental value. Most ancient philosophers – Aristotle first among them – claim that truth is a value not in itself but because it leads us to knowledge, which has a value not only because all men naturally seek it (*Met*. A, 980a22), but also because it leads to happiness or well-being as the supreme good. On this view, truth has *eudaimonic* value because it leads to knowledge and because knowledge is constitutive of well-being and happiness (Hazlett 2013). Thus truth would have only an instrumental value because knowledge is the primary value, to which truth is attached, in the sense that knowledge has more value than true belief. This view is reinforced by a famous argument in recent epistemology, the so-called "Swamping Argument" (Kvanvig 2003). When one wants to go to Larissa, and with respect to that specific goal, having a true belief about the road to Larissa seems to be just as good and valuable as knowing the road to Larissa. Knowledge is thus swamped by true belief with respect to its value (here utility). Knowledge, however, as Plato noted (*Meno* 147b), is firmer and stronger than true belief, and for this reason, better and more valuable than true belief. If we accept this claim (although we shall see below a reason to qualify the idea that knowledge is always more valuable), true belief has a value, but this value is less than that of knowledge, hence not final. But whether it is truth or knowledge which carries the load of value, their value is the value of utility either in the narrow sense or in the wide sense of promoting well-being. The question is: to what extent has truth such an eudaimonic value?

To borrow Hazlett's (2013) useful distinctions, knowledge and truth can have eudaimonic value: a) normally (in most cases), b) generally (in all cases), c) typically (in typical cases). This value can be either instrumental to well-being (when well-being is not constituted by know-

ledge) or constitutive of well-being (when it is of the essence of well-being to be constituted by knowledge), which can be either desire independent or desire independent. Hazlett formulates the *eudaimonic ideal of true belief*:

> *For any subject S and p normally believing what is true about p is better than believing what is false about p*

To say that true belief is *normally* better than false belief is meant to avoid the easy objection that there can be cases where a true belief, or indeed a piece of knowledge, can be in some sense disvaluable. Cases abound, from the weak tennis player who would be better off not believing that she is going to lose her match rather than keeping the heartening belief that she is able to win, to the garden variety cases of rational self-deception (the spouse who prefers to ignore the lipstick on her husband's collar). Nobody would deny that "sometimes the value of truth is outweighted by other considerations" (Horwich 2006). In that respect, true beliefs may be only *pro tanto* valuable. "Valuable" here means: with respect to its contribution to the well-being of the agents who have them.

The problem, however, is that true belief or knowledge is not only sometimes disvaluable and false belief or ignorance valuable, but that they could also normally be so. Hazlett (2013: ch.2) argues, mobilizing a lot of evidence from cognitive and social psychology, that self-knowledge is not only sometimes, but actually very often, a bad thing, and ignorance of one's exact credentials can be a good thing. When people indulge in systematic self-esteem, and self-enhancement biases, when they nourish false hopes, are irrealistically optimist or entertain illusions of control over their plans or their lives, they not only sometimes but most often end up better off, happier and less depressed. This involve various form of self-deception or of wishful thinking, but this is all to the good for the individual. "Don't worry, be happy". Hazlett further argues that *partiality* and *charity* biases, by which we trust our friends and lovers sometimes against evidence or display systematic confidence in what they say, not only enhance well-being,

but are also positive virtues constitutive of it. Emerson praised the value of "self-reliance". People care for other things than true belief (non-alethic goods) and there are cases where false belief is associated with non-alethic goods. Hazlett concludes that "there is no clearly identifiable pattern of cases where true belief is better than false belief". In any case, true belief seems, with respect to false belief, to enjoy no privilege and to have a quite neutral status with respect to their respective contribution to our well-being.

One might, however, wonder whether such biases are really constitutive of well-being. In the first place, it is hard to believe that well-being could normally depend upon lying to oneself or upon self-deception. In the second place, true belief may be useful to life *in general*, simply because it is necessary for action. If we take up a classical line of thinking that has been formulated most clearly by Frank Ramsey, true belief is required for successful action, and we act on the basis of our beliefs about how we could realize our desires. In this sense true belief *always* has instrumental value, just in virtue of the nature of action. Paul Horwich develops this line of thought in order to argue that true beliefs are always valuable because they lead to action:

> Directly action-guiding beliefs of the form, 'If I perform A, then X will occur'. It will clearly benefit me if I have many such beliefs and if they are all true. Because when I want a given thing and believe that a certain action will result in my getting it, then, very often, I will perform that action. And in that case, if my belief is true, this desire will be satisfied; whereas if it isn't true no such result is ensured. So true beliefs of the directly action-guiding form will indeed tend to benefit me. And the more such true beliefs I have the broader the spectrum of desires that will be easy for me to satisfy in this way. Moreover, these special beliefs are the results of inferences that tend to preserve truth; so it will benefit me for the premises of those inferences to be true. And there is no proposition that might not someday serve as such a premise. Therefore it will indeed be good for me — at least, that's what it's reasonable for me to suppose — if I believe every true proposition and if every proposition I believe is true. (Horwich 2006, 350)

This general instrumental value of true belief is independent from the occasional disvalue of some true beliefs and from the sometimes valuable nature of false beliefs and biases, and is not threatened by these exceptions.

Neither is it threatened by the familiar examples of trivial and useless true beliefs which are often adduced against the positive TG (i) version of the goal of having true beliefs. Indeed counting the number of blades of grass in the garden or of grains of sands on the beach, trying to know how many people have a name beginning with the letter "D" in Wichita, Texas, or asking oneself whether *Joe di Maggio had a 56-game hitting streak*, are idle attempts believings or knowings. Other alleged counterexamples include beliefs about things which are so esoteric that no one would care to acquire them. Now as soon as we try to specify criteria for what kind of knowledge or belief is significant or potentially significant, we run into trouble. Some very idle or trivial beliefs might turn out to be significant in one circumstance or other, whether or not we can figure out how they can be such, and valuable for one reason or another. As soon as one attends to the particular cases, there is always room for either granting these beliefs value or disvalue. But along to what axis or criterion of evaluation? It is obvious that *typically* any true belief, as idle, trivial or useless it can be, is valuable, as the Ramsey-Horwich kind of reasoning establishes. The Ramsey-Horwich line takes the value of truth to be not a property which attaches to truth in general, but only a property which attaches to each particular truth which is a candidate for being believed. For each "action-guiding" proposition, there will a specific value in believing it, in so far as it leads to successful action. The value of truth in general is only the generalization on the list of such action-guiding propositions. But the value in question is utilitarian or success-in-action value, and one might ask whether true belief cannot be valuable in general, independently of whether it leads to successful action. For isn't it the case that any truth, however trivial or insignificant, is of *epistemic* value, in so far as it is a *truth*? (Lynch 2004, 152). Aren't truth and knowledge common goods just as water and fresh air are supposed to be common goods for

mankind (Zagzebski 2003)? Here we should remind ourselves that there are different kinds of value, and in particular not only practical values, but also epistemic ones. From the fact that true belief may be disvaluable or less valuable practically, it does not follow that it is disvaluable, *period*. In particular there is a dimension of evaluation along which true belief is *prima facie* valuable, which is epistemic evaluation. In so far as truth is the epistemic goal of inquiry, *any* truth whatsoever is epistemically valuable, including the most trivial or insignificant ones. Indeed this remark does not suppress the problem of distinguishing significant from insignificant true beliefs, but the fact that all truths are epistemically good does not mean that they are all *equally* epistemically good (Treanor 2013, Pritchard 2014, 121) We can indeed sort out those which are deep and which augment our knowledge of the world from those which are idle or shallow. But that does not prevent all truths to be, in variable degree, epistemically good.

At this point we should pause a bit to think again about the "swamping argument" alluded to above. It purports to show that knowledge is no better than true belief with respect to practical purposes. But from the fact that the practical value of true belief can swamp the practical value of knowledge, nothing follows about the *epistemic* value of true belief with respect to the epistemic value of knowledge. To use again Pritchard's terms, one should distinguish the value (or the disvalue of the epistemic) from epistemic value. And the latter is to be evaluated in terms of truth, evidence and knowledge.

Now this distinction between epistemic and practical value seems to beg the question against those who ask: "Is true belief really valuable *as such*?" For what they ask, when they point out the value of self-confidence, of trust and of various biases, they are not evaluating our beliefs from the epistemological or cognitive point of view, but also from the practical one, and their point is that *in spite of* its bad epistemic credentials, belief without evidence or false belief can turn out to be beneficial for the individual and thus contribute to his overall well-being. Pragmatists of all sorts (*e.g.* Foley 1993) are fond of telling us that there is a dimension of comparability of the epistemic and of the

practical, which makes the question "What should I believe?" *both* epistemological and practical, or perhaps neither. The ill person who knows that his belief that he will recover enhances his chances of recovering is asking a question which belongs to the two dimensions. James' alpinist who asks himself what his chances are to survive if he leaps across a dangerous mountain chasm, people who compare the advantage of believing at will over those of simply following the evidence clearly reason along the two dimensions. When we talk about the eudaimonic value of true belief we certainly evaluate it from the practical point of view, and we are obviously comparing epistemic value and practical value. But does it follow that when we engage in this sort of comparison we cease to evaluate our beliefs from the cognitive point of view of view? A wishful belief or a self-deceptive belief, a self-confident belief and an attitude of trust do not cease to be false, evidentially fragile or cognitively unreasonable when they play a positive role in our lives. Beliefs, like restaurants, can be evaluated from the standpoint of all kinds of values and normative standards. One may choose a restaurant for its food, but also for its atmosphere or for its proximity. Similarly for beliefs. They can be well-founded or not, beneficial or not to the believers, aesthetically satisfactory (dandies like to believe what is gracious or sublime) or simply preferred because they are popular (those who follow fashion or snobs like to believe what the rulers of fashion or of opinion dictate). But does it mean that there is no *primary* dimension of assessment of belief? No. Wishful thinking, as useful as it can be for ostriches or for men, is always *prima facie* wrong. False, fragile or biased beliefs too. The same is true for restaurants, which have to be evaluated for their food first: one can like a restaurant for its atmosphere, but if the food is bad there is something definitely wrong. In that respect belief cannot fall short of being evaluated epistemically. As Bernard Williams (2002) reminds us: falsity is a fatal defect for a belief.

Another way of expressing the same idea is to say that epistemic evaluation is *exclusive* for belief. And here, for reasons which I am going to give in the next paragraph, it

is more appropriate to talk in terms of *reasons* rather that in terms of value. In a number of circumstances, we evaluate our beliefs on the basis of other criteria than epistemic: we have plenty of reasons to want to believe certain things. But our reasons for wanting to believe are not the same as our reasons to believe. The former are much more diverse than the latter. When believing is – in the cases when we have the power to acquire a belief – the object of a deliberation lead to an action, the reasons which we can have to want to have this belief are much more diverse than those that we have for believing period. One can want to believe something because one finds it pleasant, comfortable, beautiful, useful, and in some sense good or valuable. But however good it can be to want to believe something – and this goodness can be appreciated along many dimensions – there is only one kind of reason for believing proper: epistemic reasons, that is truth and evidence, which are "the right kind of reason" (Millar 2003, Hieronymi 2005, Parfit 2011: appendix A, Engel 2013b).

The exclusivity of epistemic reasons does not entail that belief cannot be evaluated along the other dimensions. Truth can be good or bad in quite a number of respects. It can be bad for personal life, but also good for social and political life and for democracy in general (Lynch 2004). Nietzschean thinkers can well tell us that truth as a general goal and value is a mythology in the service of power, and that worshiping truth can be in many ways dangerous. Pragmatists can well tell us that truth is useless or disvaluable in many ways, and that other social goals and values, such as solidarity. But one might wonder whether the potential disvalue of truth can be appreciated without attending to the central role of truth in *any* evaluation – good or bad – about the value of our beliefs – in other words how we could say anything about the value of truth without taking into account its *epistemic* value first. One might wonder also whether anything of value in personal or in social life could be achieved without it (Williams 2002). Truth is at least valuable *by default*: life would be much harder without it. This does not mean that it automatically adds value to life when it comes in.

## *3. The essentialist view: teleology*

The *prima facie* epistemic nature of the evaluation of beliefs suggests that the goal of believing truths and only truth is not only a goal or an aim which could be a source of value, but that it is *essentially* so, and that the relation between belief and truth is in some sense constitutive. But in what sense?

According to what we can call *essentialism* about belief, the nature of epistemic evaluation derives from the nature of belief (Hazlett 2013: ch 5, see also Fassio 2012). There are, however, different forms of belief-essentialism, which can vary along two dimensions, which are respectively the kind of state that belief is and the kind of evaluation which is appropriate. One can, on the one hand, claim that belief has an essential nature because of (i) its *metaphysical* nature as a mental state, or because (ii) our *concept* of belief has certain constitutive *a priori* features. One can, on the other hand, evaluate beliefs by using different sorts of normative notions: (iii) one can assess beliefs as good or bad, as valuable or not, along an *axiological* dimension, or (iv) on can assess beliefs as being correct or incorrect, along a *normative* or *deontic* dimension. In general essentialist approaches of the metaphysical kind (i) are associated with the axiological dimension, because belief is supposed to aim, by nature, towards a certain kind of goal, hence to have a certain kind of *teleology*. Essentialist approaches of the conceptual kind (iv) are most often associated with the normative concepts of what one *ought* to believe, or of what one has most *reason* to believe.

Teleological views have in common the idea that it is the essence of belief to have a certain kind of aim, or goal, or direction or objective. The notion of an aim or goal suggests that aiming at true beliefs is the conscious or intentional objective of the believer. Some views are intentionalist in this sense. Thus Velleman (2000) holds that the distinguishing essence of belief with respect to other cognitive attitudes (such as guessing, imagining or supposing) is to be the attitude which is such that the believer aims at accepting its content as true if only if it is true, and Sosa (2011) compares believing with the intentional

activity of an archer who tries to hit a target – here truth – and succeeds or not to reach this target. On such views, beliefs are, to a certain extent, active states of mind, and to a certain extent, kinds of actions or at least display a certain amount of epistemic agency. But it need not be so. Hume famously held that beliefs and desires have "distinct existences": belief or "reason" is concerned with only what is true or false, whereas desires are concerned with what we aim at. Anscombe (1958) and Searle (1989) have elaborated this distinction as that between two "directions of fit" – mind to world and world to mind – which respectively belief and desires have as mental states independently of whether we take them to have that direction and intend to exploit it. On this Humean view, beliefs are essentially directed at truth whether we desire or intend it or not. This is perfectly compatible with Hume's argument that belief is not a matter of the will and is an involuntary mental state. There are two variations upon the Humean view. One is the functionalist approached, pioneered by Ramsey, according to which belief is the kind of mental state which is such that it can be, together with desires, the cause of our actions. A functionalist theory of belief says that it is of the essence of belief to be the kind of state which receives input information from the environment and which, on the basis of desires, leads to behavioral outputs The other is the Darwinian approach, according to which not only belief is that very kind of functional state, but also such that natural selection has selected it as the kind of state that it is (Millikan 1990, Dretske 2001, Papineau 1999). On the teleological view, the only normativity which is involved in epistemic evaluation is the one which is attached to the value that the agent – which can be nature itself – sets on having true belief.

One can raise at least four objections against teleological view, both in its intentionalist and in its nonintentionalist versions.

The first objection concerns the intentionalist version: believing is not, at least primarily, an intentional activity. Even if it is in *some* specific kinds of believing – those which involve mental action, judgment and acceptance towards a certain content, this can hardly be true in general. One

must leave room for unintentional kinds of believing – the paradigm example being beliefs based on perception and cases of unconscious belief formation – unless one espouses the implausible view that believing always involves some intentional activity – ranging from trying to reach its truth goal to doxastic control and commitment (Engel 2005).

The second problem affects both the intentionalist and the non-intentionalist version. It has to do with the fact that whatever the goal of belief can be – whether it aims at truth, or at knowledge, or at securing our well-being – this goal can in principle be balanced against other goals and changed. If one conceives of the aim of belief in intentional terms and as a goal, one has to accept the idea that the goal can be compared with others, and that it could change depending upon the aims of the believer. The problem is that belief is hardly a goal directed activity in this sense, and quite unlike an action. When one aims at something either in the sense of intending to do it or in the sense of having a long term plan, one typically can balance this objective or this goal against at least another one. But belief is quite unlike that. Believing is not like guessing, when one hesitates between various options, since there is actually no other choice than holding true or holding false, hence adopting an epistemic stance *anyway*. Even suspension of judgment, which comes close to having the choice between taking one option or another, cannot occur between choosing between an epistemic aim and a practical aim. When one believes there is no way to balance the truth goal against a practical goal. Truth is the only goal here is, and in this sense it cannot be a goal: epistemically there is no other choice, and as we saw above, when there is an apparent choice between an epistemic a non epistemic goal, the epistemic one is always the one which imposes itself by default.

The third problem about the teleological view is specific to its Darwinian or biological version. If belief is the kind of mental state that it is, with its specific direction of fit, and its aim toward truth in virtue of its having been selected by natural selection, how can it be the *essence*, in the metaphysical sense, of belief to have these characteristics?

It is a merely contingent feature of our psychology. Can't we conceive of a distinct state – let us call it *schmelief* – which would be such that in most cases it would be directed at truth, but in other cases it could be directed at falsity, and which in any case would not invariably be directed at truth (Papineau 2013). On the Darwinian view there is no obstacle to such a supposition, and for that reason it cannot be part of the *essence* of belief that it aims at truth. There is a tension in the Humean view and in the Darwinian view. On the one hand they say that aiming at truth is a general fact about belief, which is its essence. On the other hand these views tell us that this fact holds naturalistically, hence is contingent. Hence there is no essential aiming at truth in belief, no metaphysical nature of belief.

The fourth problem has to do with the normative force of the epistemic evaluation for teleological theories. On the Humean view there is no other normativity than the direction of fit of belief. Beliefs are the kind of attitudes that have the mind to world direction of fit. But no normative advice, even less a normative prescription or guidance can be involved in this bare fact: form the fact that my beliefs are *supposed* to be true or false in virtue of being the kind of natural mental state that they are, it follows nothing about what I *ought* to believe or not, which is normally what one can expect from a normative guidance (Dretske 2001). The intentionalist version does not fare better. If the normative force of truth in believing is that of an intention to reach, through believing, the goal of reaching a truth, then this force is no stronger than that of a hypothetical imperative of the kind: *if one wants to have true beliefs one ought to acquire the belief that p*. But this imperative, being conditional on the desire or intention of the believer, is much too weak to capture the normative force of the evaluation, which is that of a *categorical* and *unconditional* imperative:

(TO) *one ought to believe that p if and only if p*

which does not depend on the condition that the believer wants or intends to believe the truth. The normative force

cannot be simply instrumental, as all the Humean views presuppose (Kelly 2003). It depends not on a prior mental state but simply on how things are for the believer. Even on the teleological reading, the aim of belief should not depend on contingent desires or intentions. It should be a fixed aim. But how can it be, given that aims can, by nature, change?

### *4. The normative account*

These objections lead us to favour an account according to which it is an *a priori* and constitutive property of our *concept* of belief that it is subject to a norm of correctness, which is truth. The constitutive correctness norm of belief, on such a view is the following

(TN) Necessarily S's belief that p is correct if and only if p, and incorrect otherwise

(TN) is supposed to be necessary, hence to treat the norm of belief as an essential feature of belief. In this sense, belief has a normative essence (Wedgwood 2007). (TO) is but one way of interpreting (TN). But (TN) can also be conceived as a conceptual *a priori* truth about belief (Boghossian 2003).

What speaks in favour of the correctness account is that, unlike the teleological one it captures the normative force of the relation between belief and truth. (TN) is not contingent upon the desires or the intentions of the believer. It accounts better (in the sense of inference to the best explanation) for some of the most pervasive features of belief. First, the fact that belief is involuntary and not under direct control of the will: if there is a norm to believe truths and only truths, any willful believing has to violate that norm. This is not to say that it cannot occur, but that if it does it has to occur against this norm. Second, the fact that "Moorean" beliefs or assertion of the form "P but I believe that not P" are paradoxical: if there were not a direct relation between one's asserting or believing that P and one's believing that P *is true* such Moorean assertions would not be strange. Third, the correctness norm for

belief accounts for the "transparency" of belief: when one deliberates about whether to believe that P, the question is settled as soon as one realizes that P is the case (Shah 2003). Remember Moore's remark quoted above: "I cannot at any given moment distinguish what is true from what I believe". If there were not this direct connection between believing that P and believing that P is true, the remark would make no sense. Fourth, the normative account gives the best explanation for the centrality of belief among other belief-like attitudes and quasi-doxastic states, such as suppositions, acceptances, guesses, imaginings, partial beliefs, tacit beliefs, subdoxastic states, creedal feelings, feelings of knowing, pathological beliefs, phobias, "aliefs", delusions, biases. Some of these attitudes and states (like guessing, imagining or supposing) resemble belief in having propositional contents and being truth-evaluable. Others are "strange bedfellows" for belief, since it is not clear that they have a propositional content or are truth evaluable. The best criterion to distinguish these from beliefs is to see whether they are subject to the correctness norm and the transparency test. And there are reasons to think that they do not pass this test.

This conceptualist account is Kantian in spirit because it involves the element of reflection: the standard or norm of right belief applies to us as soon as we reflect upon the nature of correct belief, which exists regardless whether one wants or not to conform to those standards (Hazlett 2013, 206). This reflective element is most clear in the transparency feature mentioned in the previous paragraph. One can further argue that the norms of belief, as well as those of action, form part of the normative order in the objective sense, as parts of the domain of reason (Skorupski 2011). In this sense the Kantian conception is essentialist. But it need not be essentialist in an objective metaphysical sense, in which the domain of reasons and of norms would be a domain of facts, alongside natural facts. The Kantian view is rather constructivist: it does not say that there is a normative domain of facts, among which normative facts, and among which normative intentional facts, feature. On this latter view one can conceive, and in a more Platonist and cognitivist vein, of the normative nature of belief as

an essential property of belief, and more widely of intentional states in general and take them to be (Wedgwood 2007, Parfit 2011).

There are a number of objections against the normative account of belief, which I cannot examine here. A number have to do with the specific form which the biconditional (TN) is supposed to take in order to be able to guide properly belief formation (Bykvist and Hattiangadi 2007, Gibbons 2013, Chan 2013). The main objection is that it is not clear that the norm (TN) is normative at all. On the one hand, it is supposed to prescribe what one ought to believe it is much too strong: if (TN) were the correctness norm for belief, it would necessarily motivate us to believe that P if and only if P is true. But it need not do so. The norm does not inescapably motivate us to believe. As a prescriptive norm, it is implausible (Steglish-Pedersen 2006, Glüer and Wikforss 2009). On the other hand, if it is supposed to be prescriptive about what one ought to believe, it is much too weak and it has no normative force. From the fact that it is correct to believe that P if and only if P, nothing at all seems to follow for how one has to go about with respect to believing that P. That these objections contradict each other shows that there must be something wrong with the premiss from which they start. They both presuppose that the correctness norm is normative in the sense that it ought to be *prescriptive* of our believing and guide our belief formation. But the presupposition is wrong. (TN) does not say, and doesn't have to say what kinds of beliefs we have to adopt or how. It not prescriptive in the sense of what J.J. Thomson (2008) calls "directives", and prescribes no action be they epistemic or not. It just says what we ought to believe (Engel 2013, 2013 a). One can here compare the epistemic norm of truth with what Parfit (2011, 417 sq.) says about normative truths in ethics. They are not supposed to tell us what to do or to motivate us for certain actions. They are supposed to tell us what we *ought to do*, and what kinds of reasons we have. As Parfit says, if there were no such truths about our reasons, we could not begin to ask ourselves what kinds of decisions to take or how to live. Similarly, the correctness norm for belief tells us what we ought to believe, and what

kinds of reasons we have. Such reasons do not depend upon our desires or upon our attitudes. They are objective.

Another strong objection against the normative account, especially in the cognitive essentialist sense (but also in the Kantian conceptualist form) is that it does not account for the supervenience of the normative properties or concepts upon the natural ones. The dilemma here is familiar: either the normative properties do not supervene on the natural ones and are left dangling without any natural basis. I cannot here deal with this objection. But one must remark here that any attempt to reconcile the normative essence of belief with natural facts will have at some point to assume that the norms of correct belief, and the objective reasons that there are to believe, have to depend in some sense from our psychological states, and most upon our desires. Only these can belong to the natural basis of our reasons, and only these can properly motivate us to accept the epistemic norms and to conform to them. If one takes this line (which is the one taken by most anti-realists and non-cognitivists about epistemic norms and values, especially expressivists) then one will have to reject two of the claims which I have taken to be central to epistemic norms: their categorical, non instrumental character on the one hand, and their exclusivity, the fact that epistemic reasons are by essence the "right kind" of reasons. One way or another we shall have to take exception to the supervenience of the normative on the natural.

### *5. Farewell to Plumpton*

If the foregoing rambling thoughts are correct, there is no specific problem of the value of truth, because truth is not, primarily and constitutively, a value. One can ask whether it has value, including final and intrinsic value, but any appreciation of the value of truth will have to start from an appreciation of epistemic value in general. Epistemic value is best thought of not in terms of value, but in terms of norms. The normative stance has priority over the evaluative stance. Belief is subject to epistemic evaluation first. This does not prevent us from asking whether true belief can have a practical value or contribute to well-being.

I have not given any argument here in favour of non cognitivist and realist conceptions of epistemic normativity, as against anti-realist, non cognitivist and expressivist views. But if the considerations proposed in §3 against the teleological conception of epistemic normativity are correct, they favour a realist account. It remains to be seen whether it should take the form of a buck-passing account, of a Kantian constructivism, or of some form of Platonism about norms and reasons. In all this we shall very probably have to say goodbye to Ramsey's pragmatism.[1]

## *References*


Boghossian, P. A. (2003), The Normativity of Content, *Philosophical Issues*, 13 (1), 31–45, repr. in Boghossian, *Content and justification*, Oxford, Oxford University Press.

Bykvist, K. and Hattiangadi, A. (2007), Does Thought Imply Ought? *Analysis* Vol. 67, No. 296, 277-285.

Chan, T. ed. (2013), *The Aim of Belief*, Oxford: Oxford University Press.

Clifford, W.K. (1879), The Ethics of Belief, in *Lectures and other essays*, vol.2, London: Macmillan.

David, M. (2012), How to take Truth as a Goal? In C. Jäger and W. Löffler, eds., *Epistemology: Contexts, Values, Disagreements; Proceedings of the 34th International Ludwig Wittgenstein Symposium*, Frankfurt: Ontos Verlag 2012: 203–214.

Dretske, F. (2001), Norms, History, and the Mental, in *Perception, Knowledge and Belief,* Cambridge, Cambridge University Press.

Engel, P. (2004), Truth and the Aim of Belief, in D.Gillies, ed. Laws and Models in Science. London: King's College Publications.

Engel, P. (2013), In Defense of Normativism about the Aim of Belief, in Chan 2013.


---

1. It is a pleasure to contribute to my good friend Nils-Eric Sahlin's Festschrift for his sixtieth birthday. Through him, I was introduced to a wealth of topics: Ramsey, probability theory, theories of truth, epistemology, and I have always admired the combination of formal rigor and of humanism with which he has approached these topics. Through him, I was introduced to Lund philosophy, and have not since then ceased to be inspired by the (then) Kungshuset philosophers, whom I thank for their hospitality and their inspiring community.

Engel, P. (2013a), Doxastic correctness, Proceedings of The Aristotelian Society, Supplementary volume.
Engel, P. (2013b), Beief and the right Kind of Reasons, Teorema, 32, 3, 47–360.
Engel, P. *To appear*, Can the deflationist Theory of truth Account for the Normativity of Truth?", in D. Achiouriotti, J. Martinez and H. Galinon, eds *Unifying the Philosophy of Truth*, Springer.
Fassio, D. (2012), *Belief and Correctness*, dissertation, University of Geneva.
Foley, R. (1993), *Working without a Net* ,Oxford: Oxford University Press.
Gibbons, J. (2013), *The Norm of Belief*, Oxford: Oxford University Press.
Glüer, K. & Wikforss A. (2009), Against Content Normativity, *Mind*, 118, 469, 31–68.
Hazlett, A. (2013), *A Luxury of the Understanding, on the value of true belief*, Oxford, Oxford university Press.
Hieronymi, P. (2005), The wrong kind of reason. *The Journal of Philosophy,* 102, 437–457.
Horwich, P. (2006), The Value of Truth, Nous, 40 (2), 347–360.
James, W. (1897), The Will to Believe in *The Will to Believe and Other Essays*.
Kelly, T. (2003), Epistemic rationality and instrumental rationality: a critique, *Philosophy and Phenomenological Research* , LXVI, 3, 163–196.
Kornblith, H. (1993), Epistemic Normativity, *Synthese*, 94(1993), 357–376.
Kvanvig (2003), *The value of knowledge and the pursuit of Understanding*, Cambridge: Cambridge University Press.
Lynch, M. (2004), *True to Life*, Cambridge Mass: MIT Press.
Millar, A. (2006), *Understanding People*, Oxford: Oxford University Press.
Millikan, R. (1990), Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox, *The Philosophical Review* 99(3), pp. 32–353, repr. in *White Queen Psychology*, Cambridge Mass: MIT press.
Moore, G.E. (1903), *Principia Ethica*, Cambridge: Cambridge University Press.
Mulligan, K. (2009), Values, in R. LePoidevin, P. Simons, A. McGonigal and R. Cameron, eds. *The Routledge Companion to Metaphysics*, London: Routledge, 401–412.
Papineau, D. (1999), Normativity and Judgment" *Aristotelian Society Supp*. Vol. 73, Issue 1, 16–43.
Papineau, D. (2013), There are no Norms of Belief, in Chan 2013, 64-79.
Parfit, D. (2011), *On what Matters*, vol. I and II Oxford: Oxford University Press.
Pritchard, D. (2014), Truth as the Fundamental Epistemic Good,

in J. Matheson and R. Vitz, eds, *The Ethics of Belief*, Oxford, Oxford University Press.
Ramsey, F.P. (1991), *Philosophical Papers*, ed. Mellor Cambridge, Cambridge University Press.
Ramsey, F.P. (1990), *On Truth*, ed. Majer and Rescher, Dordrecht Reidel.
Sahlin, N.E. (1990), *The Philosophy of Frank Ramsey*, Cambridge: Cambridge University Press.
Skorupski, J. (2011), *The Domain of* Reasons, Oxford: Oxford University Press.
Sosa, E. (2011), *Knowing Full well*, Princeton, Princeton University Press.
Steglish-Petersen A. (2006), The Aim of Belief: no Norm Needed, *Philosophical Quarterly* 56, 225, 500–516.
Treanor, N. (2014), Trivial Truths and the Aim of Inquiry, *Philosophy and Phenomenological Research* Volume 89, 3, p. 552–559.
Velleman, D. (2000), On the Aim of Belief, in his *The possibility of Practical Reason*, Oxford: Oxford University Press.
Wedgwood, R. (2002), The Aim of Belief, *Philosophical Perspectives*, 16, 267–297.
Wedgwood, R. (2007), *The Nature of Normativity*, Oxford: Oxford University Press.
Williams, B. (2002), *Truth and truthfulness,* Princeton, Princeton University Press.
Wright, C. (1992), *Truth and Objectivity*: Oxford: Oxford University Press.
Zagzebski, L. (2003), The Search for the Source of Epistemic Good, *Metaphilosophy*, 34:1/2 pp. 12-28; repr. in *Moral and Epistemic Virtues*, D. Pritchard and M. Brady, eds, Blackwell, 2003, 13–28.

# Ambiguity in decision making and the fear of being fooled

PETER GÄRDENFORS

## 1. *Ambiguous decisions*

In economics, decision theory has been dominated by the classical model based on maximising expected utility (MEU). The preference structures that generate choice based on expected utilities were axiomatised by von Neumann and Morgenstern (1944) and Savage (1954) (see Gärdenfors and Sahlin (1988) for a presentation of the classical theory). A central assumption behind these axiomatisations is that the decision maker can construct lotteries where two or more alternatives are mixed according to some probability distribution. The representation theorems show that if the preferences fulfil the proposed axioms, then there exist a unique probability distribution and a utility function (unique up to linear transformations) so that choices generated from the preferences can be determined by MEU. MEU has become the hallmark of Homo oeconomicus as a decision maker and it has been built into many types of game-theoretic analyses.

The decision theory based on MEU has been extremely influential in economic theory. However, there are some indomitable examples that have caused problems for the traditional theory. One is Allais' (1953) paradox that has been the subject of extensive research. Another is Ellsberg's (1961) paradox that will be the focus of this article. This paradox strongly suggests that the estimated probabilities of the events (together with the utilities of the outcome) are not sufficient to determine the decision, but the amount of information underlying the probability

estimates is also important. In Ellsberg's terminology, the ambiguity of the probabilities influences decisions.

Ellsberg (1961, pp. 653–654) asks us to consider the following decision problem. Imagine an urn known to contain 30 red balls and 60 black and yellow balls, the latter in unknown proportion. One ball is to be drawn at random from the urn. In the first situation you are asked to choose between two alternatives A and B. If you choose A you will receive $100 if a red ball is drawn and nothing if a black or yellow ball is drawn. If you choose B you will receive $100 if a black ball is drawn, otherwise nothing. In the second situation you are asked to choose, under the same circumstances, between the two alternatives C and D. If you choose C you will receive $100 if a red or a yellow ball is drawn, otherwise nothing and if you choose D you will receive $100 if a black or yellow ball is drawn, otherwise nothing. This decision problem is shown in the following decision matrix.

| | Red | Black | Yellow |
|---|---|---|---|
| A | $100 | $0 | $0 |
| B | $0 | $100 | $0 |
| C | $100 | $0 | $100 |
| D | $0 | $100 | $100 |

The most frequent pattern of response to these two decision situations is that A is preferred to B and D is preferred to C. It is easy to show that this decision pattern violates MEU. As Ellsberg notes, this preference pattern violates Savage's (1954) 'sure thing principle', which requires that the preference ordering between A and B be the same as the ordering between C and D:

> The sure-thing principle: The choice between two alternatives must be unaffected by the value of outcomes corresponding to states for which both alternatives have the same outcome.

The rationale for the preferences exhibited in the Ellsberg paradox seems to be that there is a difference between the quality of knowledge we have about the states. We know that the proportions of red balls in the urn is one third,

whereas we are uncertain about the proportion of black balls (it can be anything between zero and two thirds).

Decisions are made under risk when there is a known probability distribution over the outcomes, such as when playing roulette, and under uncertainty (ambiguity) when the available knowledge is not sufficient to single out a unique probability distribution. The problem of decision making under uncertainty has been known since Keynes (1921) who writes of the "weight of evidence" in addition to probabilities (see also Knight 1921). He argues that the weight and not only probabilities should influence decisions, but he never presents a model. Savage's (1954) axioms cannot handle this form of uncertainty. Ellsberg's (1961) paradox brought the problems of not distinguishing between risk and uncertainty out in the light and it has generated an immense literature not only in economics, but also in psychology and philosophy (for an extensive review see Etner et al. 2012). Several solutions to the paradox were proposed (e.g. Smith 1961, Anscombe and Aumann 1963, Gärdenfors and Sahlin 1982, Einhorn and Hogarth 1985, Wakker 1986), more or less following Wald's (1950) maximin rule.

I briefly summarize the solution proposed in Gärdenfors and Sahlin (1982). The first step consists in restricting the set P of all probability measures to a set of measures with a 'satisfactory' degree of epistemic reliability. The intuition here is that in a given decision situation, certain probability distributions over the states of nature, albeit possible given the knowledge of the decision maker, are not considered as serious possibilities. The decision maker selects a desired level $\rho_o$ of epistemic reliability and only those probability distributions in P that pass this ρ-level are included in the restricted set of distributions $P/\rho_o$, but not the others. For each alternative ai and each probability distribution P in $P/\rho_o$ the expected utility $e_{ik}$ is computed in the ordinary way. The minimal expected utility of an alternative $a_i$, relative to a set $P/\rho_o$, is then determined, this being defined as the lowest of these expected utilities $e_{ik}$. Finally the decision is made according to the following rule (cf. Gärdenfors (1979), p. 16).

> The maximin criterion for expected utilities (MMEU): The alternative with the largest minimal expected utility ought to be chosen.

Despite all the attempts to formulate new decision rules, it was Schmeidler who, in two ground-breaking papers from 1989 (Schmeidler 1989, Gilboa and Schmeidler 1989) solved the problem of providing a new axiomatisation containing a proper weakening of Savage's sure-thing-principle that could explain Ellsberg's paradox and some other empirical problems for Savage's model. Schmeidler incorporated uncertainty (ambiguity) aversion – as opposed to risk aversion – within a formal framework that encompasses both risk and uncertainty. In brief, he showed how to model attitudes towards risk and uncertainty directly through sensitivity towards uncertainty, rather than the indirect classical modelling through sensitivity towards outcomes (utility).

Let an "act" be a map from states of nature to the set of outcomes that a decision maker cares about, let $\leq$ be the decision maker's preference relation, and let x be any number strictly between 0 and 1 (an objective probability). Savage's sure-thing principle can then be formulated as follows:

> For any acts f, g and h, if $f \leq g$, then $x \cdot f + (1-x) \cdot h \leq x \cdot g + (1-x) \cdot h$.

Schmeidler (1989) replaces this axiom by what he calls co-monotonic independence. A simpler and even weaker condition, called certainty-independence is used by Gilboa and Schmeidler (1989):

> For any acts f and g and any constant act h, if $f \leq g$, then $x \cdot f + (1-x) \cdot h \leq x \cdot g + (1-x) \cdot h$.

It is easy to show that this condition is not violated by Ellsberg's paradox. On the basis of co-monotonic independence, Schmeidler (1989) proves a central representation theorem involving Choquet integrals. Gilboa and Schmeidler (1989) then prove a representation theorem that says that certainty-independence together with some

other more standard axioms are satisfied if and only if the preference ordering is generated by the MMEU rule defined over a convex class of probability distributions. The interpretation is that the uncertainty of the agent is reflected by the fact that the knowledge available to the agent is not sufficient to identify a unique subjective probability function but only a (convex) class of such functions.

## 2. *Fear of being fooled*

The main reason why von Neumann and Morgenstern (1944), and later Nash (1950) and Savage (1954), introduced lotteries as part of the strategy sets seems to be that this generates a convex set of alternatives that allows them to apply certain mathematical techniques in order to prove appropriate representation theorems. For example, by using probability mixtures of strategies, Nash (1950) is able to apply Kakutani's fix-point theorem to show the existence of a (pure or mixed) Nash equilibrium in all finite games.

The assumption about lotteries is, however, not very realistic from an evolutionary or a cognitive point of view. Nature is uncertain, but it almost never plays lotteries with well-defined probabilities. In other words, decision problems under risk occur mainly in ordinary lotteries, parlour games and in experiments performed by behavioural economists. Gärdenfors and Sahlin (1982, p. 364) write:

> Within strict Bayesianism it is assumed that these beliefs can be represented by a single probability measure defined over the states of nature. This assumption is very strong since it amounts to the agent having complete information in the sense that he is certain of the probabilities of the possible states of nature. The assumption is unrealistic, since it is almost only in mathematical games with coins and dice that the agent has such complete information, while in most cases of practical interest the agent has only partial information about the states of nature.

A similar point is made by Morris (1997, p. 236) who writes that according to MEU the decision maker

> . . . should be prepared to assign a probability to any event and accept a bet either way on the outcome of that event at odds actuarially fair given his probability of that event. Yet both introspection and some experimentation suggest that most people are prepared to do so only if they know the true probability.

This means that if the goal is to model human decision making, the focus should be on decisions under uncertainty. Uncertainty can be seen as having two sources: Internal when the state of knowledge is incomplete (ambiguity) and external when it is due to a chance event (Kahneman and Tversky 1982).

Furthermore, it seems that most human decisions depend on the actions of others, that is, decision theory should be seen as a branch of game theory. On the other hand, the typical game-theoretical models with well-defined sets of strategies and mappings from the choices of the players to outcomes often do not correspond to realistic decision or game situations. In real life, it is often unclear who the potential opponents or collaborators are and which the decision alternatives are. Thus the traditional division into decision theory and game theory oversimplifies the problems that are found in everyday decision making. Consequentially, it would be more appropriate to focus on the amount of information available to the decision maker and how that relates to the evaluation of the alternatives.

Curley et al. (1986) present five psychological hypotheses for why ambiguity aversion exists. Their experimental results best support the "other-evaluation hypothesis" that the decision maker "perceives as most justifiable to others who will evaluate the decision". I here propose yet another hypothesis: One factor that, in my opinion, has not been sufficiently emphasized in decision or game theory is the decision maker's fear of being fooled. The decision maker almost always has limited information about the state of the world and about the knowledge of others. She thus runs a risk that somebody knows more about the decision or game situation and that this can be exploited to the disadvantage of the decision maker (Morris 1997). For example, the one controlling the urn in an

Ellsberg type of decision situation may know that people, in general, have a preference for selecting red balls and rig the urns accordingly. Some rudiments of this hypothesis can be found in Gärdenfors (1979).

Avoiding being fooled leads to a cautious decision strategy. The proposed decision rule is: Select the alternative that maximises expected utility under the condition that the decision maker will not risk being fooled. This is an adapted version of the previous MMEU rule.

In this context, it should be noted that the general motivation for why a player should strive for a Nash equilibrium can be interpreted as avoiding being fooled. In an equilibrium no player can exploit the choices of the others.

Let me now apply the proposed decision rule to a variation of Ellsberg problem. The decision maker has a choice between two urns: (A) A risky urn that contains 50 black and 50 white balls and where she wins if a black ball is drawn. (B) An ambiguous urn that contains either (1) 25 black and 75 white balls or (2) 50 black and 50 white balls or (3) 75 black and 25 white balls. Also for the ambiguous urn she wins if a black ball is drawn.

The game promises a chance of winning $100 and no risk of losing anything. The possible gain in the game must, however, be paid by somebody (call him Somebody) and it is natural to assume that Somebody wants to minimize his losses. From the perspective of the decision maker, Somebody may know more about the distribution of the urns or may manipulate the distribution. An obvious question for the decision maker is why Somebody should offer such a bet.

For simplicity, let us assume that the decision maker believes that before she makes the choice, Somebody can, to some extent, manipulate the probability that a particular distribution is selected for the ambiguous urn. It is in the self-interest of Somebody to maximize the probability that the urn containing 25 black balls (urn 1) is selected. Assume that he can select a probability $a > 1/3$ for urn 1, and probabilities $b < 1/3$ for urn 2 and $c < 1/3$ for urn 3 (where $a+b+c=1$). The expected value of choosing the ambiguous alternative for the decision maker would then be $(0.25 \cdot a + 0.5 \cdot b + 0.75 \cdot c) \cdot \$100$. This value will be smaller

than 0.5·$100, which is the expected value of choosing the risky urn. According to the proposed rule, the decision maker should therefore choose the risky urn. This is in accordance with the empirical observations concerning Ellsberg problems.

The problem can also be described as a sort of game where Somebody's decision alternatives are how to manipulate the three urns. Let us further simplify the problem by assuming that Somebody (who is initially endowed with $100) has full control of which of the ambiguous urns is chosen. Then we obtain the following game matrix:

| | Somebody | | |
|---|---|---|---|
| | Urn 1 | Urn 2 | Urn 3 |
| Risky | $50,$50 | $50,$50 | $50,$50 |
| Ambiguous | $25,$75 | $50,$50 | $75,$25 |

It is obvious that <Risky, Urn 1> is the only Nash equilibrium. This result holds as soon as Somebody has the smallest chance of manipulating the number of balls (and wants to maximize his outcome). The upshot is that, from the game perspective, ambiguity aversion is a Nash equilibrium.

The point of this little exercise with a problem of the Ellsberg type is that the decision maker realizes that she has limited information about the ambiguous urn and that further information might lead her to change her decision. In particular, Somebody may already have further information and exploit this to the disadvantage of the decision maker. Fear of being fooled then leads her to select the MMEU solution. This is in accordance with Morris' (1997) interpretation: "Thus all we need to argue, in order to rationalize apparent uncertainty aversion in betting as a consequence of private information, is that our individual assign some probability to the person he is betting against knowing more about the true probability than he does".

### *3. Empirical evidence*

There is further empirical evidence that supports the hypothesis presented here. Firstly, if it is made clear that the

process of choosing which urn is the actual one in the ambiguous case is random (and not made by a human), the ambiguity aversion of subjects disappear. In such a situation, subjects apparently treat the ambiguous alternative more or less as the alternative with the risky urn.

Secondly, Brun and Teigen (1990) performed three experiments where subjects were asked to guess the outcome of an event such as the winner of a football match or the sex of a child. The subjects were asked whether they would prefer to bet (a) in a situation before the event or (b) after the event had taken place but where they did not know the outcome. A large majority of the subjects preferred to guess before the event had taken place. Furthermore, most subjects expressed that predictions are more exciting than postdiction and failing a postdiction causes more discomfort than failing a prediction. Brun and Teigen (1990, p. 17) speculate that "internal uncertainty is felt most acceptable when matched by a corresponding external uncertainty, and most aversive when contrasted to an externally established fact". Their experiments clearly support the hypothesis concerning fear of being fooled.

Thirdly, Curley et al. (1986) found that the ambiguity aversion increased when there was a potential for negative evaluation by others. In one of their experiments, subjects either acted (a) under a condition that their choice would become public or (b) under the condition that their choice would never be known. The result was that subjects made ambiguity-avoiding choices more often under condition (a). The subjects seem to believe that when information becomes available after a decision is made, they are judged as if they should have known the outcome, even if it was not available at the time of the decision. In general, people seem to be more hesitant to make decisions when they are missing information that will become available at a later time.

### *4. Discussion: Relation to other theories*

The literature on ambiguous decision making and the Ellsberg problem is extensive. I conclude by comparing my proposal to some other theories. Firstly, as I have already

noted, the fear of being fooled hypothesis is related to, but more specific than, the other-evaluation hypothesis of Curley et al. (1986). Secondly, analysing ambiguous decision making in terms of fear of being fooled is similar the proposal by Frisch and Baron (1988). They define ambiguity as "the subjective experience of missing information relevant to a prediction" (1988, p. 152) and they note that in an ambiguous situation "there is a possibility that an opponent will know more than you, and therefore will have an advantage" (p. 153). However, neither Curley et al. (1986), nor Frisch and Baron (1988) present any model of how the experience of missing information leads to a decision rule.

The theory that comes closest to the one presented here is that of Morris (1997). He notices (1997, p. 236) that if the decision maker does not know the objective probability of an event, there is some potentially valuable information that she does not have. But then he focuses on betting situations where the very fact that Somebody offers a bet that the decision maker finds favourable is an indication that Somebody has some relevant information that the decision maker does not possess. As Ramsey (1931) writes: "The old-fashioned way of measuring a person's beliefs is to propose a bet and see what are the lowest odds which he will accept. This method I regard as fundamentally sound. ... [But] the proposal of a bet may inevitably alter his state of opinion". The main bulk of Morris's paper is devoted to an analysis of such betting situations. In contrast, my focus has been on decision making in general that do not only involve betting situations. For me, the fear of being fooled is a general constraint on decision making.

Another approach to ambiguity aversion that also focuses on the information dynamics of the decision process is the anticipated regret theory of Krähmer and Stone (2013). The following quote summarizes their proposal:

> The agent cares about both his material payoff and his ex post evaluation of the performance of his action relative to a reference action—the ideal action that he believes he should have chosen had he known with foresight what he knows with hindsight. If the actual payoff falls short of his

> reference action's expected payoff, he experiences regret and his utility falls; otherwise, he rejoices and his utility rises. (2013, p. 713)

The regret they consider occurs when the decision maker realizes that he should have chosen otherwise. A central part of their argument is that when, for example, the outcome of a drawing from an ambiguous urn is revealed, it contains information about the contents of the urn. A consequence is that the outcome may then make him regret his choice. In contrast, the outcome from a risky urn does not give the decision maker any new information and so gives no reason to change the choice.

The main difference between my hypothesis concerning fear of being fooled and that of Krähmer and Stone (2013) is that I involve a hypothetical Somebody who may manipulate the ambiguous urn, while their theory only concerns the thoughts and potential regrets of the decision maker. In principle, it is empirically testable which interpretation of decision makers' deliberations best explains their behaviour when faced with a situation of the Ellsberg type. However, two theoretical arguments already now speak in favour of my hypothesis. Firstly, Krähmer and Stone's proposal involves rather sophisticated self-reflection in terms of potential regret on part of the decision maker, while my hypothesis is only based on everyday reasoning concerning individuals with conflicting interests (both wanting to gain as much as possible). Secondly, Bovens and Rabinowicz (2015) criticize Krähmer and Stone's (2013) account of regret. In brief, they claim that "it is a mistake to think that in comparisons of potential regret and potential joy the reference point for regret and for joy coincide".

To sum up, I have presented a hypothesis based on fear of being fooled as an explanation of subjects' behaviour in situations of ambiguous decision making. I have presented some empirical evidence in favour of the hypothesis, but future testing will decide its fate. I also believe that the fear of being fooled can explain other aspects of human decision making, but that topic will be reserved for a later occasion.

*Acknowledgement*

I am grateful for helpful comments from Jörgen Weibull. I also thank the Swedish Research Council for its support to the Linnaeus environment Thinking in Time: Cognition, Communication and Learning.

*References*

Allais, M. (1963) "Le comportement de l'homme rationel devant le risque, critique des postulates et axioms de l'école americaine", *Econometrica* 21, 503–546.

Anscombe, F. J. and Aumann, R. J. (1963) "A definition of subjective probability", *Annals of Mathematical Statistics* 34, 199–205.

Bovens., L and Rabinowicz, W. (2015) "The meaning of 'Darn it!", Hirose, I. and Reisner. A. (eds.), *Weighing and Reasoning: A Festschrift for John Broome*, Oxford: Oxford University Press.

Brun, W. and Teigen, K. H. (1990) "Prediciton and postdiction preferences in guessing", *Journal of Behavioral Decision Making* 3, 17–28.

Curley, S. P., Yates, F. and Abrams, R. A. (1986) "Psychological sources of ambiguity avoidance", Organizational Behavior and *Human Decision Processes* 38, 230–256.

Einhorn, H. J. and Hogarth, R. M. (1985) "Ambiguity and uncertainty in probabilistic inference", *Psychological Review* 92, 433–461.

Ellsberg, D. (1961) "Risk, ambiguity, and the Savage axioms", *Quarterly Journal of Economics* 75, 643–669.

Etner, J., Jeleva, M. and Tallon, J.-M. (2012) "Decision theory under ambiguity", *Journal of Economic Surveys* 26, 234–270.

Frisch, D. and Baron, J (1988) "Ambiguity and rationality", *Journal of Behavioral Decision Making* 1, 149–157.

Gärdenfors, P. (1979) "Forecasts, decisions and uncertain probabilities", *Erkenntnis* 14, 159–181.

Gärdenfors, P. and Sahlin, N.-E. (1982) "Unreliable probabilities, risk taking and decision making", *Synthese* 53, 361–386.

Gärdenfors, P. and Sahlin, N.-E., eds. (1988) *Decision, Probability and Utility: Selected Readings*, Cambridge: Cambridge University Press.

Gilboa, I. and Schmeidler, D. (1989) "Maxmin expected utility with non-unique prior", *Journal of Mathematical Economics* 18, 141–153.

Tversky, A. and Kahneman, D. (1982) "Advances in prospect theory: Cumulative representation of uncertainty", *Journal of Risk and Uncertainty* 5, 297–323.

Keynes, J. M. (1921) *A Treatise on Probability*, London: Macmillan Co.

Knight, F. H. (1921) *Risk, Uncertainty and Profit*, Boston, MA: Houghton Mifflin.
Krähmer, D. and Stone, R. (2013) "Anticipated regret as an explanation of uncertainty aversion", *Economic Theory* 52, 709–728.
Morris, S. (1997) "Risk, uncertainty and hidden information", *Theory and Decision* 42, 235–270.
Nash, J. F. (1950) "Equilibrium points in n-person games", *Proceedings of the National Academy of Science* 36, 48–49.
von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.
Ramsey, F. P. (1931) *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, New York, NY: Humanities Press.
Savage, L. J. (1954) *The Foundations of Statistics*, New York, NY: John Wiley and Sons.
Schmeidler, D. (1989) "Subjective probability and expected utility without additivity", *Econometrica* 57, 571–587.
Smith, C. A. B. (1961) "Consistency in statistical inference and decision", *Journal of the Royal Statistical Society* Ser. B 23, 1–25.
Wakker, P. (1986) *Representations of choice situations*, Ph.D. dissertation, Department of Economics, University of Brabant, Tilburg.
Wald, A. (1950) *Statistical Decision Functions*, New York, NY: Wiley & Sons.

# NIPT: Ethical aspects

GÖRAN HERMERÉN

## *1. Approaches*

Nils-Eric Sahlin and the late Jan Wahlström contributed greatly to the work of SBU (Swedish Council on Health Technology Assessment) and SMER (Swedish National Council on Medical Ethics) by writing early about the ethical issues raised by new developments in prenatal testing and they also discussed NIPT (Non-Invasive Prenatal Testing). In a recent update of the SBU report Nils-Eric has carried on with this work. In this tribute to Nils-Eric I want to sum up and continue the discussion of these ethical issues, since the development of tests and diagnostic methods in this area is rapid.

Focus can then be on problems raised *only* by NIPT or on problems raised by NIPT but *also* by other prenatal testing methods. Both deserve in my view to be discussed and examined. The purpose of this paper is to identify concerns raised by NIPT but not exlusively by NIPT. The reason for identifying such concerns is not to create difficulties but to clarify ethical issues, make underlying value conflicts explicit, identify research needs and find ethically acceptable solutions of the concerns – which should be in the long term insterest of everyone involved.

The advantages of NIPT are well known: it is easy, only a blood sample is required. It can be carried out early in pregnancy, and it is non-invasive; thus there is no risk for miscarriage. There are several uses of NIPT, and they can be classified in two main groups: medical and non-medical. The former ones include uses of NIPT to identify

- RHD status
- Trisomies (T21, 18, 13)
- X-linked diseases

...

The latter non-medical uses of NIPT include:

- Paternity testing
- Identification of sex for social or cultural reasons

Ethical issues are related to the uses. Thus it is important to begin by identifying the uses – and the possibility that one particular (and non-controversial) use may pave the way for other (and more controversial) uses.

Why this focus on the trisomies? Why not on Fragile X, for instance? The standard answer is that politicians and the regulatory authorities will have to decide the list of what to look for. But if such a list is to be perceived as legitimate, it has to be preceded by a broad discussion of the social and ethical issues underlying the construction of the list. Here national councils and ethics committees have a particular responsibility.

It may be tempting to begin by comparing and contrasting two fairly recent statements:

> Fetal RHD detection in early pregnancy using a single-exon assay in a routine clinical setting is feasible and accurate... Both sensitivity and specificity were close to 99% provided samples were not collected before gestational week 8. (Wikman et al. Noninvasive Single-Exon Fetal *RHD* Determination in a Routine Screening Program in Early Pregnancy. *Obstet Gynecol* (2012);120:227).

The second statement is

> ... unique features of NGS (Next Generation Sequencing) and WGS (Whole Genome Sequencing), such as the amount and quality of data and the open-ended opportunities render the detailed view and magnitude of ethical issues somewhat different for WGS compared to other tools and strategies... most of the existing ethical frameworks ...were not drafted with the potential applications of NGS in mind... (Pinxten W, Howard HC. Ethical issues

raised by whole genome sequencing. *Best Practice & Research Clinical Gastroenterology* 2014;28:272).

The first of these statements calls attention to possibilities, the second to problems and ethical concerns.

Ethics is an academic discipline with its own theories, methods, bank of knowledge, chairs, conferences and periodicals. But it is also at the same time a practical activity, integrated in health care and medical research, with focus on identifying and dealing with conflicts of value. These value conflicts can in their turn be conflicts between the rights, interests, freedoms and obligations of those involved.

NIPT as well as many other methods of prenatal testing raises several types of value conflicts, for instance, (a) between the health and quality of life of the fetus and right of the woman/couple to decide ('reproductive autonomy'), (b) between the principle of equal dignity and rights of all humans and the risk that in practice eugenics is introduced via the back door, (c) between the duty of the physician to inform and the right not to know of the tested, particularly *re* unsolicited findings, (d) between cost-effectiveness and integrity, between the wish to use cost/effectively collected data for R&D (via 'data sharing') and the desire to protect the privacy and anonymity of those tested.

To clarify the nature of these conflicts we need to identify the stakeholders and those involved, and their values – what they want to achieve and avoid in the short and long term – but also make explicit

- Which of these values are conditional, based on assumptions?
- Which duties and rights are prima facie, which are absolute?
- Can/should values be ranked differently in research and clinical disgnostics, depending on use?

The relations between some of the key issues to be discussed here can be indicated as follows. The starting point is a clinical question. The accuracy of the test in focus is

relative to this question. The accuracy becomes problematic if there are unsolicited findings, and these findings will make it difficult to maintain and protect the right not to know. This right, and how it is dealt with, has implications for what the information should contain. Issues of informed consent are also relevant to ways of handling the data now and in the future. Who should have access to them, what information should be provided about this to those offered or asking for tests? Finally, there are issues of justice (who should have access to the tests) with economic and commercial implications.

## 2. *Accuracy and unsolicited findings*

Accuracy can in this context be defined in terms of sensitivity and specificity. Any test can be more or less sensitive and specific, and this may change over time. If a test is reasonably accurate (some false positives and some false negatives are difficult to avoid), it is from a clinical perspective reasonable to go on and ask, if the test is actionable or not. Is there something that could or should be done, provided that the test result shows this or that? If a test is indeed actionable we can carry on and ask whether it is actionable directly or indirectly, and by whom.

False positive test results is, of course, a problem with both medical and ethical implications; and it is essential to continue to improve the test methods. According to information I have received from Bo Jacobsson and Erik Iwarsson, it happens that persons decide not to take an invasive test to verify the test result of the non-invasive test, if the latter has been positive. This may increase the number of abortions and lead to abortions of fetuses that would anyway have resulted in miscarriage a few weeks later. The information challenges raised by this development should not be underestimated.

NIPT can be used to find out if the fetus has chromosomal aneuploidies, but unsolicited findings can also be obtained. How are they to be dealt with? And how is this possibility reflected in the information to the pregnant woman/the couple? This is not a new debate, but NGS (Next-Generation Sequencing) will increase the possibil-

ity and probability of such findings. There is a recent discussion of such issues in e g Rigter T et al. Reflecting on Earlier Experiences with Unsolicited Findings: Points to Consider for Next-Generation Sequencing and Informed Consent in Diagnostics, *Human Mutation* 2013;34:1322–28; and in Pinxten W, Howard HC. Ethical issues raised by whole genome sequencing. *Best Practice & Research Clinical Gastroenterology* 2014;28:269–79.

In the discussion of solicited and unsolicited findings, it is obviously important to distinguish between two dimensions: the probability of being affected and the severity of the condition a particular finding indicates. Needless to say, both dimensions admit of degrees, and they can vary independently of each other.

A temporary solution of the conflict between the duty of the physician to inform and the right not to know of the tested, particularly *re* unsolicited findings, has been proposed by American College of Medical Genetics and Genomics. This solution can be summarized as: "Do not offer opt out for unsolicited findings." Why not? This is difficult to understand and this proposal is also rightly criticized by SM Wolf et al in *Science* 2013:1049–50.

### *3. The right not to know*

A special problem is raised by information about increased risk for late onset diseases. WGS, whole genome sequencing, of the fetus would make this possible. In most cases it may be possible to identify a moderately increased risk of being affected by a late onset disease of varying severity (diabetes,...). But if the late onset disease is a dominantly inherited disease like Huntington, it would be possible to find out that also one of the parents may be at risk to be affected.

This will hardly be compatible with Council of Europe's Oviedo convention, art 10:2, according to which "Everyone is entitled to know any information collected about his or her health. However, the wishes of individuals not to be so informed shall be observed".

Is it possible to inform *before* the test about possible, and *after* the test about actual, findings, in such a way that

the right not to know, according to this convention, is respected? Is the right not to know conditional or unconditional, absolute? As it is stated, it is in my view unconditional; no conditions which could provide a basis for exception are mentioned.

What do tested persons want to know? What do they not want to know? This can be investigated by empirical studies. But what are they entitled to know? To answer this and the following question value premises have to be made explicit, unless the questions are understood as a request for information about what the tested indivduals are entitled to know according to currently accepted guidelines.

What should be reported to the tested person pre/post test? The possibilities include:

1. Everything that NIPT can show? (But this will be too much, and will take too much time to confirm)
2. Only part of (1): what has predictive value for the health and quality of life of the tested person?
3. (2) and/or potentially for the health and quality of life of the family of the tested person?
4. (3) but only what there is effective treatment of today?
5. only what is a direct threat against the health of the tested person unless medical interventions are made soon?
6. and what is relevant for future reproductive choices: carrier status for a autosomal recessive diseases?

This problem has been discussed by Berg et al, in Deploying whole genome sequencing in clinical practice and public health. *Genet Med* 2011;13:499–504. But these authors suggest a categorization of findings ('binning') different from the one above:

- Clinically actionable results must be reported back to the tested person
- Results of unknown or no clinical significance are not returned
- Clinically valid but not directly actionable results

are returned depending on the preference of the tested person

The binning exercise raises two problems:

1. how do you decide in practice what ends up in every bin?
2. Is there a difference between research – clinical examination?

*4. Information and consent*

The patient will have to live with the decision – and will have to take responsibility for it. This can be taken as an argument for the position that this is the patient's decision in the sense that the patient ought to be entitled to make the decision. But the decision taken will have implications for others, in particular the expected child. Anyway, the choice and the decision patients will make is highly dependent on what they know and believe. The information received thus has a key role. So:

1. What is explicitly said? By whom?
2. What is suggested? By whom?
3. What is understood by (1) and (2)? By whom?
4. How is this understood? By whom?

Is there a danger that the ease and quickness of NIPT will trivialize the decision? It has been suggested that…"practitioners will view the consent process for prenatal diagnostic testing differently depending on whether it is an invasive or non-invasive test". (Van den Heuvel AJ, et al. Will the introduction of non-invasive prenatal diagnostic testing erode informed choices? *Pat Educ Couns* 2010; 78:24–28. )

There are a number of key questions to consider at this point, such as: What should the *pre* test information contain about the testing options, the accuracy of tests and screening methods, the possibility of unsolicited findings, the right not to know, and the handling of the data collected? What should the *post* test information contain

about actual – solicited and unsolicited – findings and their accurracy? How should this information be conveyed (orally, in writing, both) and by whom? What efforts should be made to ensure that the information provided has been understood correctly? How should consent be sought and documented? These problems are not new, but they need to be discussed, particularly as the prenatal testing methods are developed rapidly.

How can these problems be dealt with? It is essential to distinguish between problems related to the number of alternatives, related to whether the methods are simple or not, invasive or non-invasive – and problems related to the difficulties to provide information about the testing and its possible results in such a way that the persons approached will understand the information and can make well-considered decisions – in their own long-range interest.

One possibility would be to introduce a two-step approach with some kind of screening or risk assessment as a first step. This would make pre/post test counselling possible – thus maintaining the same type of procedure as when invasive test methods are used. This has been proposed by e.g. Deans Z, Newson A. Should non-invasiveness change informed consent procedures for prenatal diagnosis? *Health care analysis* 2011;19:122–32). A possible drawback is that the two-step approach will appear artificial if not needed (if NIPT can replace diagnostic testing), and will then increase the costs unnecessarily.

There are also other options which could be tried in various combinations, such as letting all counselling be carried out by trained genetic counsellors, which requires considerable investment in education. But such investments may be needed anyway.

The need of education has been stressed by many, including by M Hill et al Uses of cell free fetal DNA in maternal circulation, *Best Pract & Research Clinical Obst and Gynaecology* (2012) and by Sayres LC et al Cell-free fetal DNA testing: a pilot study of obstetric healthcare provider attitudes toward clinical implementation (*Prenat Diagn* 2011; 31:1070–76), showing that health care providers in the US felt uncertain "in their current knowledge of NIPD".

## *5. Slippery slopes and criticism from disabled*

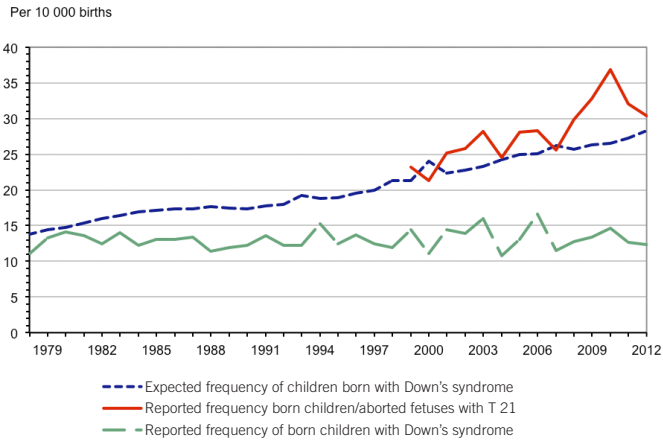
What are the benefits of early diagnostics? If the conditions identified can be cured or prevented, the benefits are obvious. But if the only treatment is termination of the pregnancy? Some advantages are clear enough, as late abortions are more risky and stressful for the woman. But there are also some dangers, which will be discussed below.

Slippery slopes and criticism from disabled is the other side of the many advantages of NIPT. NIPT can be carried out early, before the pregnancy is visible and before any bond between mother and fetus has been developed. The procedure does not, like ultra sound, make it possible for the pregnant woman to see a picture of the fetus. A blood test: simple and easy, with no risk of miscarriage – but with many consequences for those involved, especially if and when NIPT is combined with the rapid development of genome research and sequencing, NGS and WGS. This can make it easier to consent to the testing and to interrupt the pregnancy.

Brian Skotko once asked: "With new prenatal testing, will babies with Down syndrome slowly disappear?" (*Archives of disease in childhood*, 2009; 94:823–826). Even if this is not the expressed intention of carrying out the cffDNA test, it can be a not intended – but perhaps not directly unwanted – consequence of general use of the test.

This suggests that empirical information might be relevant at this point. Has the number of children with mb Down diminished – in absolute or relative numbers – as some have feared? The diagramme on the next page, obtained from Karin Källén (personal communication), provides a condensed answer to this question.

The situation is not quite the same in all Nordic countries. In Denmark it has been reported that national screening halves the number of newborns with Down's syndrome. In 2012 49 children were born with Down's syndrome in Norway, which was a record figure. The same year 135 children with mb Down were born in Sweden, according to information from Thomas Jansson. In Sweden, the number of children born with Down's syndrome remains relatively constant.




Now, if the official intention is *not* to reduce the number of children with mb Down by introducing or allowing new test methods, how are these differences to be explained? Is the reason different intentions, different policies, different information or different criteria in the Nordic countries? KUB (combined ultra-sound and biochemistry) offered to every pregnant woman in some countries but not in others?

It will clarify the situation to distinguish between problems at several different levels. At a general ethical and biopolitical level, we may ask as policy-makers or citizens: "Do all humans have the same dignity, the same rights and the same right to be born?" At the individual level, everyone can ask himself or herself: "Do I want a child with mb Down, Fragile X…?"

The point of making this distinction is simply that if the answer to the second question is *no*, this does not necessarily imply a negative answer to the first question. The reason for this negative answer can be practical, "I cannot cope with it; I already have a disabled child; the children I already have would not get enough attention etc". Perhaps the person asked could have coped with a Down child if the support of society to families with Down children had been different.

Obviously, it is one thing to want to eliminate a disease, and a very different one to want to eliminate those who have this disease. In other words, it is quite possible to

combine the effort to eliminate a disease with efforts to help those who suffer from it to live as good a life as possible.

But disabled persons and their organization sometimes do not always appreciate to be reminded about this distinction. Some may fear stigmatization and will choose NIPT and then abortion if there is a risk for chromosomal aneuploidies. Perhaps they do not want to hear: "You could have asked for NIPT. But you decided not to. Then suit yourself". In this way eugenics may be introduced by the back door.

The French National Ethics Committee (CCNE) has discussed this extensively in a report (nr 120, 13 April 2013) and stressed inter alia that the use of cffDNA (cell-free fetal DNA) should be viewed in a broader perspective and discussed against the background of ideas about disease, health, normality and toleration of variations. CCNE also stresses that large investments from society are required to make it possible for the disabled to live a good life. But which are the financial implications of this? Health economic studies of the costs and benefits of different tests are not enough.

Incidentally, the difficulties in estimating in a non-arbitrary way the quality of life of children with certain conditions, like mb Down, and their parents, should not be underestimated.

### *6. Access, costs and commercial aspects*

Who pays? If county councils will pay, there will be a priority setting problem, since resources are limited. The benefits for the patient(s) in terms of health and quality of life will somehow have to be compared to the benefits if these resources had been spent on other interventions. This raises a number of ethical issues, which have been discussed at some length (in my paper, Prioriteringar: val och värden i vården, *Årsbok 2014*, *Vetenskapssocieteten i Lund*, 2014: 65–96 with further references).

If, however, patients will have to pay out of pocket, not everybody may afford the cost of testing. This may then pave the way for a slow increase of children with particular

trisomies in certain socioeconomic groups, which raises both ethical and social concerns. Follow up studies are required to see if these concerns are justified.

Of course, the cost will also depend on whether NIPT will be offered to all pregnant women, or only to some risk group – and in that case, to which one? Here there are several alternatives, which ought to be evaluated in the light of what we know today about the testing options, the acurracy of the tests, costs and effects – as well as ethical principles. Priority setting concerns will be raised also by NIPT – as by the introduction of any diagnostics or therapy. What could we get if the resources instead were used for some other purpose?

There are several possible principles on which access could be based.

- Every pregnant woman is offered the test?
  Reason: justice, equity
- Every pregnant woman who asks for the test should be offered the test?
  Reason: fairness, equality, respect for the right not to know
- Every pregnant woman who will benefit from the test will be offered the test?
  Argument: patient needs/benefits, cost/effectiveness.
- Every pregnant woman who belongs to some risk group will be offered the test?
  Reason: the same. But which risk groups in that case?

Risk groups can be identified by using other tests, such as KUB. A decisive question will then be what these tests show, how accurate they are, which chromosomal aneuploidies are missed etc?

Which costs and whose costs are to be taken into account? The cost for the test, for the counselling, for false positives and false negatives, for indirect costs (impact on consumption or production during sick leave)? WGS will require analysis and interpretation of collected data based

on cooperation of experts from different fields. The cost of this can be considerable according to Mardis ER, The $1000 genome, the $100,000 analysis? Genome Med 2010;482:288. The cost may, however, become cheaper over time.

Suppose it is argued that the economy of a society will benefit if healthy children are born, if these children grow up and get good jobs, contribute to productivity, pay taxes etc and that children needing a personal assistent around the clock are expensive for a society. It is one thing to assert or deny such statements, and something else to say that these economic aspects should be taken into consideration when resources for health care and social services are to be distributed. The extent to which such costs are to be taken into account is not an economic issue but an ethical one – ultimately based on ideas about what kind of society we want to live in, hand over to our children and grandchildren, and to their children and grandchildren.

Strong commercial interests can directly influence the marketing of certain options and indirectly the choices persons will make. GENDIA in Belgium offers testkits via internet directly to consumers and promise 99% accuracy. The test can identify three trisomies (13, 18, 21) and X och Y in 20 ml blood from the pregnant woman for 850 euro. LifeCodexx is according to CCNE (2013:16) marketing another test, PrenaTest, for 1250 euro per test. These tests are sold without any pre/post test counselling. This will give rise to even more concern if and when the tests are developed to discover increased risks for more conditions.

There are other companies competing on this market. For instance, PrimBio is a biotech company providing NGS services to scientific communities world-wide using the latest Ion Torrent platform. In an ad which I received on August 13, 2014, the company stresses that they are 1 of 18 companies chosen by Life Technologies to be a certified Ampliseq exome service provider.

The internet market is difficult to regulate. But which requirements can, may, ought or should be implemented concerning quality control and marketing of such test kits, and on the responsibility of the manufacturers of their products and its uses?

How should collected data now and in the future be handled? Obviously, collected data must be treated confidentially in order to protect the privacy and integrity of those tested. At the same time it is reasonable to ask what can be done to promote cost/effective use of the collected data e.g. via 'data sharing'. How this is dealt with will also have implications for the information provided: what are those tested to be told about this, about how long the tests are saved, about who can have access to them, and if the samples are going to be re-analysed?

### *7. … And finally: What is new?*

Many of the issues discussed here are not new but are raised also by other methods used in prenatal testing. But this does not mean that they don't need to be examined and discussed. On the contrary, the most acute problems concern the contents and delivery of information and how to obtain a free and informed consent, standard topics in this context, and to whom NIPT should be offered.

The new issues, difficult to regulate, are raised by the rapid development in genome research and sequencing, NGS and WGS, and the possibility to combine NIPT with breakthroughs in these areas. This will open up for testing of many, more or less serious, abnormalities and conditions.

To get enough time to arrive at robust decisions, I propose that a working party with obstetricians, gynecologists, geneticists and ethicists should take a forward-looking approach and try to draft guidelines for how to deal with some of the most urgent ethical issues in this area, considering not only the current situation but what soon will be around the corner.

*(This text is based on a lecture given at the NFOG meeting 2014 in Stockholm.)*

# Roboethics: What problems should be addressed and why?

MATS JOHANSSON

## *1. Introduction*

New and emerging technologies provide ethicists with things to think about. As humanity becomes more competent and learns how to do and create new stuff, choices need to be made regarding how to use these new capabilities. Here ethics comes into the picture. The ethicists' job is rarely easy, however. For one thing, ethicists typically have to aim at a moving target. Novel technologies tend to undergo rapid development, which implies that some ethical issues may suddenly emerge, while other issues may just as suddenly turn obsolete. The ethicist's challenge becomes even tougher when merely possible future scenarios need to be addressed, something that definitely is the case in areas that are future oriented by nature, such as 'roboethics' – a discipline that Gianmarco Veruggio in 2002 characterized as follows:

> Roboethics is an applied ethics whose objective is to develop scientific/cultural/technical tools that can be shared by different social groups and beliefs. These tools aim to promote and encourage the development of Robotics for the advancement of human society and individuals, and to help preventing its misuse against humankind.

In this essay I will briefly outline some ethical considerations that ought to guide roboethicists when deciding on

what issues to explore. It should be noted that these considerations are of general relevance, and may thus, if sound, provide guidance in many other areas as well. However, their practical importance grows when there is substantial danger of future oriented speculation. There will undoubtedly be such speculations in roboethics, since that discipline is full of distractions, with promises of mindboggling and fascinating applications and innovations. The roboethicist needs to resist the temptation to engage in what is cool or fascinating rather than in what is ethically important.

## *2. Roboethics as a smorgasbord*

Until quite recently, robots were more or less equal to science fiction. Real life robots were mostly found in academic labs and on factory floors. This situation is clearly about to change. Robots have already entered many areas, ranging from health care to warfare, and they are today found sorting books, cleaning floors, and mowing lawns. Soon enough robots will be seen most everywhere. Many of us will probably live to see the day when autonomous cars bring ordinary people to work, while other robots at the same time tend to their homes. But we may also live to see the day when many people unwillingly turn unemployed due to the availability of robots capable of performing most jobs, and doing it faster, better, and cheaper than humanly possible.

What then does the future hold when it comes to robots? While it is easy enough to point out the direction, it is very difficult to predict the destination. One reason for this has to do with the rapid progress in robotics and in its supporting sciences. Another reason, I believe, is that robotics is so open-ended. This means that robotics, in theory, is consistent with most kinds of innovation. In fact, the very notion *robot* is typically defined along lines that allow for endless possibilities: artificial (mechanical or virtual) agents designed to (fully or partly) autonomously perform certain tasks.

Thus, there is a smorgasbord of areas to explore, both scientifically and ethically. The ethicist therefore needs to

decide which of these – current, future, or merely hypothetical – are worth addressing. After all, time is a scarce resource, and we therefore must use it well. We might of course identify general problems that are relevant to many possible future scenarios – killing several birds with one stone, so to speak. In the end, however, we must abandon some areas in order to explore others. And there are many decisions to be made. Should we focus on areas such as the ethics of robot prostitutes, the ethics of robotic warfare, or the ethics of robot caregivers? And what specific issues should be explored in greater detail? Consider autonomous driverless cars for example. Ought ethicists to address the "morals" of these vehicles, or who is accountable for the robots' decisions, or yet something else? Or, are the more important problems in some different area?

Merely finding a topic interesting won't do as a justification for putting a lot of effort into analyzing this topic, and it surely should not convince those who consider funding such work. Rather the ethicist should approach the decision from an ethical point of view. This is so, not because the ethicist ought to abide by higher moral standards than ordinary people, but because some academic disciplines are intimately linked to the project of making the world a better place. The latter goes for applied ethics, including roboethics. So, how to proceed?

## *3. Assessing ethical importance: seven questions*

The roboethicist's decision can be formulated in terms of what *problem* he or she ought to address. Although there is more to applied ethics – including roboethics – than problem solving, what matters at the end of the day is whether or not the ethicist comes up with something – a recommendation, an analysis, ideas, principles, or other tools – that help decision-makers (whoever those might be) to deal with problems in an ethically acceptable manner. The ethicist should, in other words, focus on what he or she can achieve, on the *outcome*.

Obviously, there are normative theories that emphasize the moral importance of other things than outcomes, such

as respect for human dignity, rights, duties, and moral character. Here however, focusing on outcomes is less committing than it might first appear, since it leaves open what should be considered morally relevant. The list of things to avoid could include physical harm, psychological suffering, human rights violations, the undermining of human dignity, etc., and worthy aims could include health, justice, happiness, people reaching their full potential, and more.

Helping decision-makers to reach the right decisions is a reasonable goal, but it provides ethicists with imprecise guidance. It remains to be shown what things the ethicist should take into consideration when deciding on what problem to address. Picture therefore a scenario [S] that in some important way involves robot(s), and that gives rise to, or exemplifies, a problem [P] that could be solved if some decision-maker reaches the right decisions. One question immediately comes to mind:

*1. What is the cost of failing to successfully deal with P, if we face P in real life?*

It arguably does not make much sense to put effort into addressing a problem if nothing of substance is at stake. And although there are possible exceptions, as when S and P are part of a thought experiment that provides valuable insights, the importance of a problem (in applied ethics, that is) is arguably related to the cost of not being able to deal with the problem in the right manner.

Typically, there is no safe route. For example, assume that we misjudge what gives robots moral standing. If so, this can, depending on the nature of our mistake, imply that we either "overprotect" machines, or that we neglect these artificial individuals' legitimate interests (or rights). Both scenarios come at great costs, either economical or moral.

So far, the cost of failure has been conditioned on the occurrence of S and P. In practice however we have to deal with uncertainties. This brings us to the second question:

*2. How probable is it that S will occur, and that as a result we will face P in real life?*

Probabilities obviously need to be taken into consideration when dealing with scenarios that are merely possible. Otherwise we would be lost in a universe of imaginable scenarios, including those found in science fiction movies as "2001", "I robot", "AI", and "Star Wars".

Needless to say, however, probabilities as such do not reveal the importance of P (other than perhaps in extreme cases where there is virtually *no* likelihood of S). If nothing is at stake, it does not matter whether or not S is likely to occur. And if much is at stake, then even a slim risk of actually having to deal with P might be enough to justify addressing P (this irrespectively of whether we seek to maximize expected utility or if we apply some kind of precautionary reasoning).

Watching out for possible (and sufficiently likely) future disasters may seem as a good idea, and 'disasters' here includes missing out on extremely beneficial outcomes. And robots have been associated with great risks. Nick Bostrom points at one possible disaster in his recent book *Superintelligence. Paths, Dangers, Strategies*. And the problem he discusses concerns how to successfully control an artificial intelligence whose capacity to reason widely surpasses ours. According to Bostrom controlling such an AI may be the last significant challenge we will ever face. If we succeed then this AI will take care of us from that point and onwards. But if we fail, we face a future in the hands of a superior being whose "basic values" differ from our own.

Before we are in a position to evaluate the importance of a problem we must take yet another aspect into consideration, namely the expected temporal distance between us (as potential problem solvers) and the problem:

*3. When (if at all) is S expected to occur?*

Time arguably matters. If we strongly suspect that S (and hence P) will not occur for a very long time, we should consider giving priority to other more urgent problems, even if the cost of failure (cf. question 1) is substantial.

Optimally we would address, and solve, those problems that are next in line. But in reality we face uncertainties concerning both *when* (if at all) the relevant situation will occur, and the *amount of time* it will take to come up with an ethically acceptable solution.

Another reason for waiting to address a problem can be that we in the future are in a better position to understand and analyze the relevant circumstances. Here it may be instructive to look back some twenty years in order to realize how fast things can change, both in society as well as in science. Who could for example have foreseen what the Internet would bring? It is one thing to predict that something will turn out to be a success or that it will alter the way we live, and it is another thing completely to correctly spell out the ways in which this something will affect our lives. It is the latter that provides us with a better picture. In roboethics such uncertainty can, for example, involve the psychological and societal impact of the introduction of social robots and robot lovers. Extrapolating from the contemporary level of technology and the society of today will perhaps do when addressing matters that concern the nearby future, but it might be very hard to comment on a more distant future where the technology has been perfected.

Time clearly introduces another dimension relevant to the roboethicist's decision, and this makes things more complicated. (It could have been much worse, however, since it really comes down to assigning probabilities of S related to specific periods in time.) And it becomes even more complex as the ethicist needs to take the *decision-maker* into consideration:

*4. Will the decision-makers be able to successfully deal with P without any help of ethicists?*

If the decision-makers will do fine without assistance the ethicist seems superfluous. The "solution" to P might be obvious, or it may coincide with what is deemed successful according to some other standard that is likely to guide decision-makers. We need no ethicist, for example, to explain that we ought to avoid malfunctioning household

robots that risk causing danger to persons or property by accidentally (or intentionally) setting fire to buildings. CEOs and their staffs of lawyers and engineers will be both motivated and capable to identify and handle this specific problem. This is not to say that other problems related to household robots can, or will, be handled correctly without ethics support. What, then, about the ethicist's contribution?

*5. Is it likely that the ethicist in question is able to, single-handedly or together with peers, make a significant contribution to the ethical analysis of P?*

Both ethics and robotics involve methods and theories that might be difficult to understand and to use. Hence, there is no guarantee that a certain roboethicist has (or can obtain) the competence level required to adequately analyze P. Strange as it initially might seem, the right thing to do to might be to get out of the way, thereby making it easier for more capable ethicists to do the job. But even then there is no guarantee that anyone will listen to reason:

*6. Will the decision-makers be guided by the ethicist's advice (on the assumption that reasonable effort is put into providing such guidance)?*

Some quests may seem rather futile. Take 'killer robots' as an example – that is military robots made to target and eliminate human enemies – and the question whether these should be prohibited. Many ethicists think that they ought to be banned altogether (although there is no consensus on this matter). And recently Clearpath Robotics became the first robotics company to take a stand against 'killer robots'. Personally I cannot picture that there will be an effective international ban on such robots. Why? There is simply too much at stake, I believe, in terms of money and power. In fact Meghan Hennessey, marketing communications manager at Clearpath Robotics, pointed (unintentionally?) at the very heart of the problem when commenting on the policy to *Business Insider*: "We're

choosing to value our ethics over potential future revenue." My guess is that other companies will target those revenues, and there will be a market. This prediction may be overly pessimistic, or fatalistic. My point remains that there may be cases where decision-makers are very unlikely to listen to reason, and that this is relevant when choosing what problems to address.

Armed with arguments (based on the answers of the six questions mentioned above) the roboethicist must now return to the smorgasbord:

*7. How important is P in relation to other problems?*

If P is regarded as less important than at least one other problem, the ethicist should think twice before putting a lot of effort in analyzing P. But of course, division of labor is a vital part of working effectively, and this means that it *might* be all right (ethically speaking) to focus on some less important problem (cf. question 5). This also relates to the individual ethicist's own expertise. It could be a waste of time and energy, all things considered, to abandon problems related to ones' own field of expertise just because there are more important problems in other areas. But this line of reasoning assumes that the ethicist focuses on important problems.

*4. Where to start? Super-problems?*

Humanity is on its own when it comes to dealing with the challenges of tomorrow (at least before the arrival of superintelligence). Unfortunately, the human species may not be very well equipped to identify, evaluate, and solve many of the important problems we will face. Or to quote Nick Bostrom:

> Far from being the smartest possible biological species, we are probably better thought of as the stupidest possible biological species capable of starting a technological civilization—a niche we filled because we got there first, not because we are in any sense optimally adapted to it.

We have set out on the quest to create autonomous machines that eventually will surpass us in every conceivable respect. And there are plenty of mistakes that can be made. The first one consists in failing to get the priorities right, and as a consequence addressing the wrong problems. We must be aware not to confuse what is merely cool or fascinating with what is ethically important. This is why the seven questions need to be addressed, again and again if needed.

I suggest that we start by making up our minds if there are any 'super-problems', i.e. problems whose importance widely exceeds that of most other problems. If we come to the conclusion that there indeed is at least one such super-problem, this has some radical implications. Waiting for philosophers and scientists to take interest in the problem, and then trust that they will do a good job, will clearly not do. Rather the appropriate strategy (by governments) is to actively recruit the most brilliant and competent persons alive, and let them analyze and solve the problem (I realize that it rules out me… but it also, by definition, most likely rules out you too). These persons should be provided with all the means required to do a good job. Whether this group is in need of ethicists remains to be seen, since that would depend on the very nature of the problem to be dealt with. But any problem that touches on how to make robots make the right decisions is likely to involve ethics.

# Ambivalenta bilder

JOHAN LASERNA

Jag känner inte en enda människa som går omkring och grunnar på det bästa och säkraste sättet att ta sig till paradiset. Inte heller någon som undrar över hur märkligt oviktig denna fråga blivit trots att den under bortåt hundra generationer var den viktigaste av alla. Så många sekler av oro och längtan efter att få veta vad som verkligen gällde för att man en dag skulle kunna få vakna upp och sträcka på benen i den evigt lummiga lunden. Vart tog drömmen om den himmelska trädgården vägen? Varför talar vi aldrig längre om den?

Annat var det på 1500-talet. Då var den rätta vägen till paradiset seklets mest angelägna och omstridda fråga. Några av dem som brottades särskilt hårt med den var Martin Luther, Jean Calvin och Huldrych Zwingli, och var och en av dem blev till sist övertygad om att vara den ende som lyckats klura ut det rätta svaret. Att deras olika svar visade sig vara helt annorlunda än påvens och den etablerade kyrkans skulle få stora religiösa och kulturella konsekvenser, inte bara för Europa utan så småningom för stora delar av världen.

Man kan undra varför. Kunde inte var och en få tro som den ville och leva därefter? Nej, om det fanns något som påven och de olika reformatorerna var helt överens om så var det att det var okristligt att låta var och en bli salig på sin fason. Det kunde rimligen bara finnas ett svar på frågan om hur man kommer till paradiset. Sanningen måste alltid vara en. Och det är Gud som sitter inne med svaret.

Var och en av dessa paradisgrubblare var förstås övertygade om att det bara var de själva som hade direkt tillgång till Guds tankar och att alla andra uppfattningar än deras

1. **SYNDAFALLET.** Från vänster till höger berättar bilden hur Eva först tar emot en frukt från kunskapens träd av en reptilkvinna och sedan erbjuder den åt Adam. Därpå förmanas de båda av Gud, som ser till att en eldröd ängel driver ut dem ur den trygga muromgärdade trädgården. Nakna och skamfyllda tvingas de ut i en karg och farlig vildmark. Ur hertig Johan av Berrys tidebok.

egen var en skymf mot honom, ett djävulens bländverk som det var varje rättrogen kristens skyldighet att bekämpa.

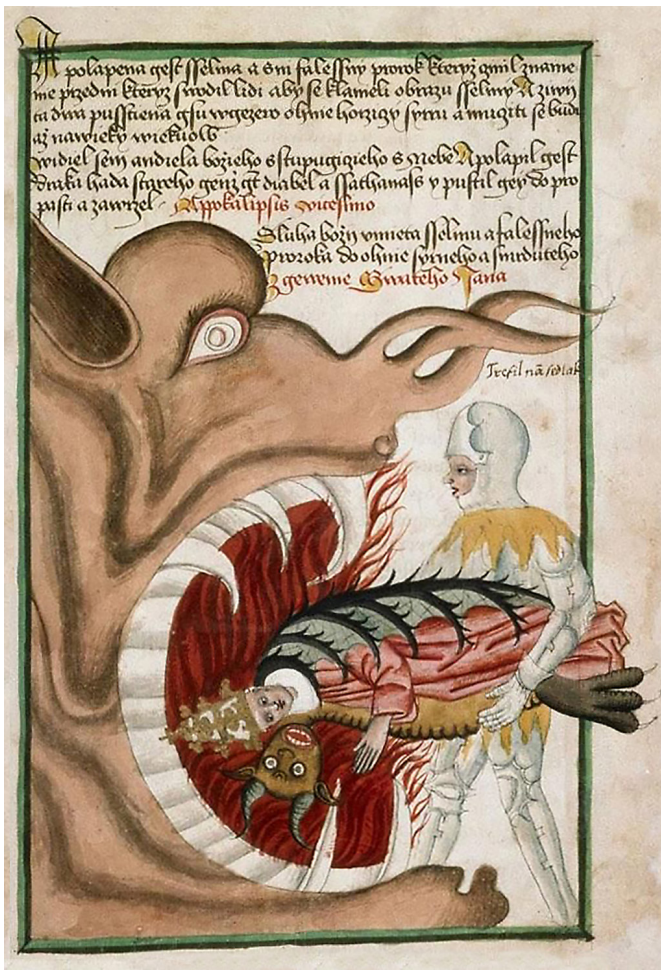
Att ha fel uppfattning i vad ens omgivning uppfattar som viktiga frågor är alltid riskabelt. Ofta farligt. Ibland bokstavligen livsfarligt. Paradisfrågan var länge en sådan fråga för många människor. Ingen gjorde sig ostraffat sin egen bild av hur människan skulle finna den snåriga stig som ledde tillbaka till den eviga trädgården. Och så kom det sig att kristna under 1500-talet började slakta andra kristna i olika religionskrig som varade i mer än hundra år.

## *Livsfarliga bilder*

Miljontals människor mördades under den här tiden. Märkligt nog inte bara för sina föreställningars skull utan även för sina bilders. Vad är det med bilder som berör oss så djupt? De tycks öppna upp något, men det är svårt att riktigt säga vad. De är bara materia och form, ändå upprättar de en sorts parallell verklighet. Som om varje bild skapar ett hål där något lockande och farligt kan sippra in. Som om bilder vore någon sorts tröskelvarelser. Eller dubbelnaturer. Ibland tycks de besitta en kraft som kan få oss att älska dem, men också att vilja slå sönder dem. Eller rentav slå ihjäl den som håller sig med fel sorts bilder, eller förhåller sig till dem på fel sätt.

Kyrkorna i västra Europa var i början av 1500-talet fyllda av just sådana kraftfulla bilder. Det kunde röra sig om väggmålningar, skulpturer eller altartavlor med änglar, helgon, kyrkliga dignitärer, bibliska personer, Jesus eller jungfru Maria. Det var viktiga bilder. Dyra att framställa och med stor makt över människors sinnen. Alltför stor tyckte reformivrarna. Calvin, Luther och de andra trosförnyarna var alla mycket medvetna om den kluvenhet som bilder skapar, deras både-och-status som ett slags revor i verkligheten. De uppfattade klart och tydligt bildens inneboende ambivalens, och de gillade den inte.

Bilder har makt att lura in människor på fel väg, menade dessa bildtvivlare. De erbjuder falska löften. De vill ersätta Gud med sig själva, med avbilder, ersätta ande med



2. **PÅVEÄTARE.** En hybrid mellan satan och påven får vad han förtjänar eftersom han erbjudit strafftidsreduktioner i skärselden i utbyte mot kyrkobyggen och avgudabilder: han langas in i helvetesgapet. Bokmålning ur ett hussitiskt manuskript från ca 1490. Jan Huss var ledare för en böhmisk reformrörelse redan under tidigt 1400-tal.

simpel materia. Den katolska kyrkans alla storslagna och påkostade bilder är förföriska, och det är just det som är det farliga. De frestar människor att dyrka själva föremålet istället för det som föremålet avbildar. En bild är bara ett redskap för att kommunicera med Gud. Gud bor inte i bilderna. Bilder är falska gudar, fulla av lögner. Bilder är förmätna, listiga och svekfulla. Att vörda dem är att tillbe dem och därmed att bryta mot det andra budordet, att inte göra sig bilder av Gud. För ingen människa kan göra sig en bild av Gud. Bara Gud kan se sig själv.

Att donera pengar till kyrkomålningar och skulpturer var för reformatorerna därför knappast någon god gärning. Att dessutom tro att det skulle ge en fördelar på den andra sidan var inget mindre än hädelse. Och ruttnast av alla var den katolska kyrkans dignitärer, som med påven i spetsen bit för bit sålde ut såväl himlen som Kristi frälsningsverk i utbyte mot kyrkobyggen fulla av avgudabilder.

Allt detta ledde så småningom till en massförstörelse av det kristna kulturarvet. Religiös konst blev i de reformerade områdena en självmotsägelse, något oundvikligen blasfemiskt. Ett djävulens bländverk. Att förstöra bilder blev som så många gånger förut i mänsklighetens historia en god gärning, rentav en plikt, och även ett sätt att visa sin avsky för de personer som ledde människorna vilse. Att slå sönder en altartavla med helgon, eller i bästa fall en påve, var som att ge den onde själv vad han förtjänade. Och samtidigt visa Gud att man hatade de falska gudarna.

Plundringen av kyrkorna i norra Europa, vandaliseringen av målningar, skulpturer och föremål hade på så sätt sin motsvarighet i likvideringen av dem som hade fel uppfattning om det rätta sättet att komma till himlen. Att förstöra bilder och att döda kättare gick hand i hand. När det gäller så viktiga saker som att utplåna felaktiga föreställningar får man inte lämna något åt slumpen.

Så förstördes ett rikt kulturarv, särskilt i norra Europa och i Nederländerna, där den religiösa konsten under senmedeltid och renässans utvecklats till så sublima nivåer. Kyrkorna tömdes. Skulpturerna slogs sönder. Bilderna brändes. På sina ställen blev bara de tomma väggarna kvar. Den rena arkitekturen i Guds rensade tempel.

3. **IKONOKLASMENS BILDER.** Biskop Guy van Avesnes vandaliserade gravskulptur i katedralen i Utrecht. 1500-talets våldsamma bildstormare skapade med sina bestraffningsliknande aktioner mot skulpturer och målningar i många fall nya och ofta starkare bilder än de som lämnades orörda.

*Bilder hjälper oss att leva*

I de katolska delarna av Europa var man inte lika bildskeptisk. Vid konciliet i Trient fastslogs att bilder av Jesus, den heliga jungfrun och helgonen skulle finnas kvar i kyrkorna och att de skulle vördas men inte tillbes. Man tillber och dyrkar bara Gud, menade konciliet, inte bilderna. Dem vördar man, på samma sätt som man vördar helgon, reliker och änglar, det vill säga att man hyser aktning och respekt för dem. Att inte göra det är att inte vara kristen.

Det är svårt att inte se något klokt i detta förhållningssätt till bilder. Ikonoklasmen känns på något sätt onaturlig, hur obekvämt det ordet än känns i munnen. Vårt behov av bilder sitter djupt. Vi är utpräglat visuella varelser. Intelligenta, men också känslostyrda, lättlurade, vidskepliga och lättskrämda. Vi behöver bilder som berör oss, som talar till oss på andra vägar än genom språket, bilder som vi delar med andra, som för oss närmare varandra, som gör världen och den mänskliga samvaron begripligare för oss. Bilder hjälper oss att leva. Vi vet att de bara är bilder. Ändå tar många av oss dem för mycket mer än så.

Någon gång i det djupa förflutna började människor göra märken, tecken och föremål med betydelser. Och sedan dess har vi ägnat oss att försöka förstå dem. Bilderna utvidgade den mänskliga repertoaren. Men de skapade också en ny form av verklighet. Och nya slags frågor och uppgifter. Vad var det bilderna ville säga? Vad ville de berätta? De flesta bilder som människor skapat, såväl nu som då, berättar på språk vi inte förstår. Vi ser dem. Urskiljer saker i dem. Kanske ser vi två personer och några föremål på en bild. Vi kan då enkelt ge en uttömmande beskrivning av den. Men vi vet egentligen inte vad bilden vill. Eller om den vill något alls.

*Bilder i träd*

Så länge jag kan minnas har jag sett bilder i träd, framför allt ögon och ansikten, men också kroppsdelar som armar, ben, torsor och kön. Varje dag går jag en promenad med min hund och passerar en liten backe med hundraåriga ekar. Och varje dag ser jag nya fingrar, munnar och kropps-

öppningar i dem. Överallt skjuter näsor och bröst ut. Eller monstruöst förvridna gap. Ge mig en stund med ett träd och snart är det förvandlat till ett pseudokubistiskt jättekollage av ihoptotade deformerade kroppsdelar. Som en sorts naturens egen psykiska automatism. Jag kan inte säga var detta smått maniska antropomorfa bildskapande kommer ifrån, men jag tror knappast att jag är ensam om det. Människor verkar ha en läggning för att se mänskliga former i naturen. I moln, stenar och klippformationer. Eller i precis vad som helst, som rostade brödskivor eller kaffeskum.

Leonardo da Vinci hade vetat vad jag talar om. I ett råd till en bildkonstnär skrev han:

> Betrakta vilken vägg som helst som har fläckar eller som är gjord av olika sorters sten. Om du är i färd med att framställa en scen kommer du att i väggen kunna se något som liknar olika sorters landskap med berg, floder, klippor, träd, slätter, djupa dalar och kullformationer. Du kommer också att kunna se strider här och där, figurer i hastig rörelse, ansikten med märkliga uttryck, egendomliga klädedräkter och en oändlig mängd ting som du sedan kan förenkla till fristående former.

Några av hans teckningar är precis så där drömskt otydliga, som fuktfläckar på en vägg. Man anar sig till föremål och levande varelser i dem men mycket förblir osäkert och olika betraktare ser olika saker i samma fläckar. En del bildskapare verkar medvetet ha utnyttjat den här egenheten som många av oss har till att skapa kryptobilder, bilder som inuti sig själva gömmer andra bilder. Kunskapens träd utformas ibland just så, som både träd och något annat, ofta ett människoskelett, och varslar på så sätt om det öde som väntar den som äter av dess frukt.

Sådana moraliserande kryptobilder har förstås en alldeles speciell lockelse, upptagna som vi alltid är med att skilja vän från fiende och pålitlig från opålitlig. Och de lämpar sig särskilt väl i propagandasyfte, där de effektivt avslöjar motståndarens rätta ansikte.

4. **IKONOFILERNAS RÄTTA ANSIKTE?** På en underligt formad kulle bedrivs alla möjliga former av bildtillbedjan. På slänten nedanför är det motsatsen som gäller. Här slår man glatt sönder och eldar upp det man dyrkar uppe på kullen. Samtidigt bildar landskapet och gyttret av människor det kadaverliknande huvudet av en munk, med flint och allt. *Allegori över ikonoklasmen*, etsning av Markus Gheeraerts den äldre, 1568.

*Hatshepsut*

De gamla egypterna var lika besatta av bilder som vi och 1500-talsmänniskorna. Lika övertygade om deras förmåga att öppna portar till andra världar. Och beredda att slå sönder dem när det behövdes.

Hatshepsut var drottning i det egyptiska riket för ungefär 3500 år sedan. Hon tillhörde den 18:e dynastin och lyckades med något som ingen annan egyptisk drottning före henne hade gjort. Hon avancerade till farao. I hennes gravtempel Djeser-djeseru i Deir el-Bahri, nära Luxor, finns monumentala väggreliefer som pedagogiskt redogör för hur guden Amun förklädd till Hatshepsuts far gör drottning Ahmose gravid. Stenen talar sitt tydliga och ovedersägliga språk: det är guden själv som sett till att placera Hatshepsut på tronen.

Hennes efterträdare Thutmosis III hade en helt annan åsikt. Efter några framgångsrika militära kampanjer, som utvidgade det egyptiska riket till ett historiskt maximum, övergick han till att bedriva en samvetsgrann historierevision på hemmaplan. Han såg till att Hatshepsuts namn ströks ur alla officiella register över rikets faraoner och skickade ut hantverkare att eliminera alla avbilder man kunde finna av henne i form av statyer och offentliga väggreliefer. Det är svårt att avgöra skälet till vendettan. En ledtråd kan finnas i att det bara var bilder som visade Hatshepsut som farao som förstördes, inte de som visade henne som drottning. Kanske var det ett sätt för Thutmosis III att säkra sin egen sons legitimitet. Genom att undanröja rivaliserande successionslinjer tillbaka till gudarna, till Osiris och Horus, banade han väg för Amenhotep II:s uppstigande på den egyptiska tronen.

Thutmosis selektiva ikonoklasm skulle få kreativa efterföljare. Hundra år senare blev Aeknaton farao och försökte föra fram solskivan Aten som den ende guden. Ett hopplöst företag kan tyckas, i ett rike med tusentals gudar. Samtidigt var det just farao som ytterst finansierade alla gudatempel i riket. Att strypa stödet till dem hade en förödande effekt. Instruerad av den nye ende guden, Aten, drog Aeknaton och hans hustru Nefertite ut i öknen och lät bygga en helt ny huvudstad vid nuvarande el Amarna,

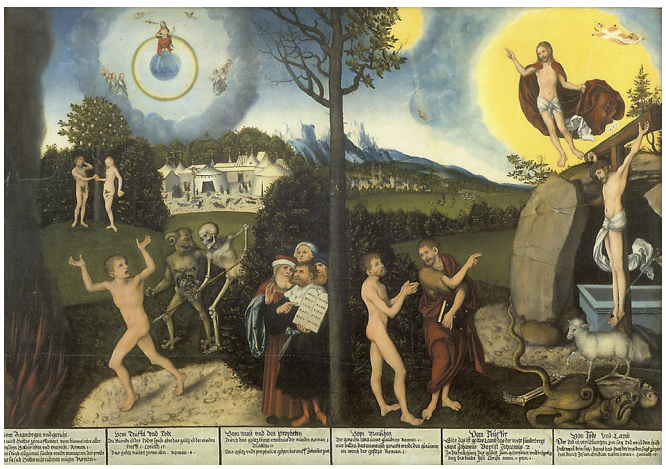
5. **HATSHEPSUTVÅLDNADER.** Ikonoklasternas mejslar har efterlämnat märkliga silhuetter och ordgåtor på väggarna i Karnaktemplet i Thebe. Andra utplåningstekniker gick ut på att systematiskt hugga bort hela rektangulära fält eller att hugga ner reliefen helt och jämna till ytan med ett lager gips och sedan hugga in ett nytt bildelement. Ännu mer drastiska metoder var att plocka ut hela block och ersätta dem med andra eller att dölja Hatshepsutfigurena genom sätta upp nya stenväggar framför dem. Det mest radikala ingreppet var att riva hela helgedomen.

där de införde sin monolatriska statsreligion. Men experimentet blev impopulärt, både bland de som levde i den nya huvudstaden och hos den övriga befolkningen. Egyptierna var inte alls benägna att överge sina många gudar. Deras olika kulter levde kvar i rikets städer och tempel. Aeknaton lät därför sända ut nya kadrer av stenhuggarikonoklaster, den här gången för att helt enkelt förstöra så många som möjligt av konkurrerande gudars avbilder, namn och tempel.

Vad som kallats den första totalitära staten kollapsade snart och en turbulent period på några decennier följde, där tre faraoner snabbt avlöste varandra. Snart kom Horemheb, nästa bildstormare, till makten. Även han förstod sig på bilders makt och sände genast ut stenhuggare i alla riktningar. Men Horemheb var ingen simpel imitatör. Istället uppfann han en egen och förfinad form av bildstormning. Stenhuggarnas uppgift den här gången blev inte att som tidigare brutalt hacka bort misshagliga bilder och tecken från offentliga byggnader och tempel utan att ändra om alla avbilder av de fyra föregående faraonerna så att de istället liknade Horemheb. Dessutom skulle alla förekomster av deras namn på stenar och papyrus ersättas med hans eget. På så sätt sträckte han elegant ut sin egen regeringstid trettio år tillbaka i tiden. Ett briljant statskonstgrepp som förmodligen inte ens Stalin skulle ha kunnat gå i land med om han hade kommit på idén att försöka.

### *Den tämjda bilden*

Konsten att tämja bilder har alltså gamla anor. Och den har ständigt utvecklats. Till sist insåg även Luther att man istället för att förbjuda bilder kunde använda dem för att undervisa om den sanna vägen till himlen. Bara man såg till att de inte kunde missförstås. Det gällde att kuva bilderna. Att domesticera dem. Få dem att gå i den rättrognes ledband. Att utnyttja bildernas kraft och makt till att undervisa människorna om de rätta sakerna. Säkrast gjorde man det genom att inte bara, som förut, återge scener ur Bibeln, utan genom att skapa ett slags pedagogiska visualiseringar av den korrekta teologin och förse dem med



6. **BILDER I STRYPKOPPEL.** Lucas Chranachs komprimering av den lutherska teologin, utförd enligt Luthers egna instruktioner. Bildens vänstra del försöker visa att den lagbundna kristendomen inte kan rädda människan från syndafallets konsekvenser. Den leder istället till helvetets lågor. Den högra bilden visar den enda räddningen: tron på den återuppståndne Jesus ofattbara nåd. Trädet i mitten grönskar instruktivt bara på nådasidan. Längst ner summeras bilden i sex olika teman, alla belysta med citat ut Nya Testamentet. Målningen är daterad 1529.

relevanta bibelcitat. Bilderna blev bibelutläggningar, konfession, predikan. Och strängt underordnade texten.

Men man undrar om det fungerade. Även de mest programmatiska och textförsedda bilderna från den tidiga reformationen innehåller tvetydigheter och detaljer som hotar att sätta det teologiska systemet i gungning. Det är verkligen inte lätt att utläsa en entydig och klar berättelse ens ur den lutherska reformationens mest domesticerade målningar, som till exempel *Lagen och nåden*, av Lucas Cranach den äldre. Och inte blir man klokare av att konsultera Luthers egna utläggningar av de olika bibelcitaten som placerats under bilden.

Även dessa reducerade målningar, renskrapade från postbibliska inslag, innehåller kanske lite för mycket. Bilder är lika svåra att tämja som betraktarna av dem. Och de väcker oftast långt fler frågor än de ger svar. Så var det förmodligen för de människor som såg reformations bilder när de var nya, men kanske i än högre grad för oss som möter dem sent på jorden, nära femhundra år efter att de kom till, och som lever i en kristen kultursfär som blivit så annorlunda, som har genomgått lutherdomens bildreningsbad och lärt sig finna ro och skönhet i vita väggar, och som inte längre tycks särskilt angelägen om att ha ihjäl den som har en annan föreställning om hur man bäst försäkrar sig om en plats i himlen.

Men även vi förhåller oss till bilder som om de betyder något. Som om de förtjänade vår vördnad och respekt. De förtrollar oss. Vi vill skydda dem, äga dem, förstå dem. Och precis som många andra före oss vändas vi i vår ambivalens för dem. De är ju bara bilder. Ändå finns det något i oss som vill tro att de är något mer än så. Det fick mästerförfalskaren Hans van Meegeren erfara när han 1945 arresterades för att ha sålt målningar av Vermeer till utlänningar, bland annat Hermann Göring. Han undgick dödstraff genom att inför rätten visa att det var han själv som hade målat de så kallade mästerverken. Ingen säljer ostraffat ut sitt lands konstnärliga nationalklenoder. Inte heller gör man sig lustig på vissa profeters bekostnad. Det kan än i dag kosta en livet.

## *Bilder från paradiset*

Bilder är uppenbarligen farliga. De förvrider våra huvuden. Vi har orimliga förväntningar på dem. Samtidigt är de underbara. Vi kan inte leva utan dem. Vissa av oss lever helt och hållet för att skapa dem. Andra för att äga dem. Ytterligare andra för att förstå dem. Men även om vi varken är konstnärer, konstsamlare eller konstvetare så undgår ingen av oss att påverkas av dem. Och hur vi än förhåller oss till dem kan vi inte låta bli att värdera dem. Så gjorde även Federico Borromeo, den katolske teologen och ärkebiskopen i Milano som var övertygad om att Gud skapat världen genom färger. Naturen är en färgpalett. Skapelsen är ett konstverk. Gud är den främste av alla målare. Ungefär så resonerade Borromeo, och blev tidigt i livet besatt av bilder. De blev för honom ett annat sätt att närma sig Gud än genom Skriften. En bild kan ge betraktaren en försmak av himlen. En bild kan vara porten till paradiset. Konsten är en väg till Gud. Han började samla på bilder. Och knyta till sig de främsta av det sena 1500-talets målare i Italien och Flandern. En av dem var Jan Brueghel den äldre.

Under sitt långa liv i bildernas våld målade Brueghel över fyrtio motiv med en blombukett i en vas. Var och en av dessa målningar tog flera månader av hårt arbete att färdigställa. Tröttnade han aldrig? Vad är det med alla dessa snarlika bilder av blommor som fångade honom så? I ett brev till sin mecenat Borromeo skrev Brueghel den 14 april 1606:

> Jag har för Ers Nåds räkning påbörjat en bukett blommor som kommer att bli mycket vacker, såväl för sin naturtrogenhets skull som för sin sällsynthet. En del av dem är antingen helt okända eller sällan påträffade häromkring. Jag har därför varit i Bryssel för att direkt från naturen avbilda några blommor som inte går att uppbringa här i Antwerpen.

Fyra månader senare skrev han i ett annat brev till Borromeo att bilden med buketten skulle komma att innehålla över hundra olika sorters blommor, alla i naturlig storlek.

7. **ORDLÖS SJÄLAVÅRD.** En av Jan Brueghel den äldres många blomstermålningar: *Blomstervas med juveler, mynt och snäckskal*, 1606.

Så många och sällsynta sorter har nog aldrig förfärdigats med lika omsorgsfull uthållighet. Denna målning kommer att bli vacker att se på under vintermånaderna. Några av färgerna ligger mycket nära naturens.

Ändå är det ingen naturalism i modern bemärkelse som de båda männen traktade efter. Bukettens generösa överflöd har ingenting med verkligheten att göra. Brueghels målning avbildar något omöjligt, kanske paradisiskt. Den är ingen övning i realism, ingen didaktisk bild avsedd att tillfredsställa en encyklopedisk vetgirighet. Istället är den en ödmjuk inlevelse i Guds skapelseakt. Eftersom alla dessa växter blommar vid helt olika tidpunkter på året och på helt olika ställen visar bilden i själva verket fram en plats utanför tid och rum. Buketten är ett utomvärldsligt kuriosakabinett, ett förnämt urval av naturens överdåd och Guds konstnärliga skaparkraft.

Dessutom fångar och förevigar bilden varje växt i dess flyktiga höjdpunkt, i själva blomningsögonblicket. I den värld som målningen visar fram finns inget vissnande, ingen död. Allt levande där är ungt, friskt, fruktsamt.

Även själva arrangemanget strider mot naturens lagar. Ingen mänsklig florist skulle gå i land med uppgiften att ordna denna illusionistiska blomstervägg, denna tapet av precist placerade fyrverkerier av färg och form, där ingen enda skymmer någon annan, där var och en smeks av ett milt och varsamt ljus, evigt lysande i sin strikta hierarki, med övervägande små blommor nertill och sedan allt fler större ju högre upp buketten reser sig.

För Borromeo och Brueghel var en sådan bild ingen avbild, inget man kastade ett öga på för att få en stunds förströelse eller hängde på väggen för att visa upp sin raffinerade smak. Den var mycket mer än så. Den gjorde det möjligt för betraktaren att vårda sin själ. Den var ett redskap för att i meditation närma sig Gud genom hans verk. En sådan bild erbjöd en väg till paradiset.

### *Kampen mellan ordet och bilden*

Brueghels blomstervas och Cranachs iscensättning av Luthers doktrin vill alltså leda oss till samma mål, men på helt olika vägar. Hos Cranach är bilden teologins lydiga

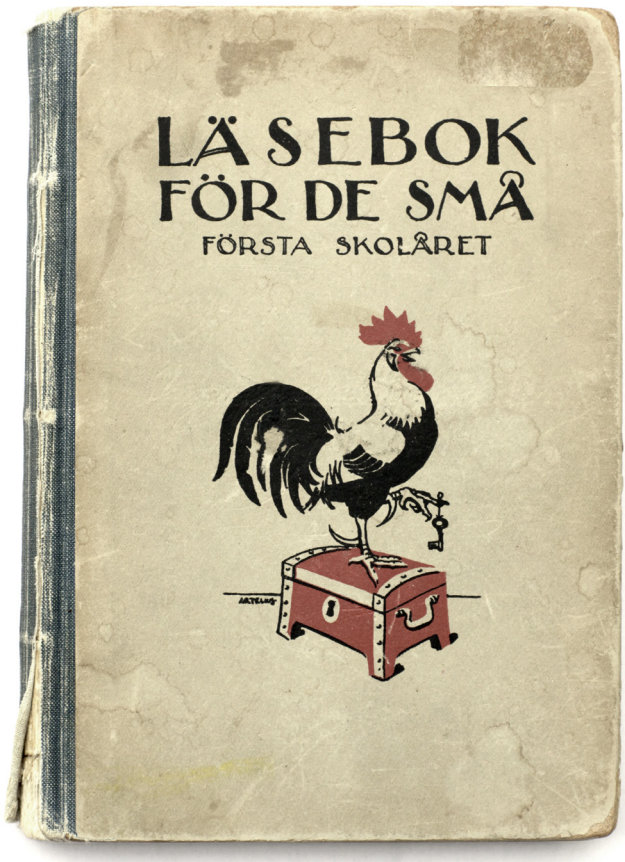
dragdjur. Bibelns ord håller i piskan. Hos Brueghel är allt bild. Inga ord sufflerar dess mening. Färgerna och formerna talar naturens eget övernaturliga och ordlösa språk.

Den här kampen mellan ordet och bilden ser man om och om igen i den europeiska kulturhistorien. Perioder av bildutveckling avlöses av ikonoklastiska reningsbad. Men bilderna kommer alltid tillbaka. De smyger sig in från obevakade hörn, frodas i utkanter och marginaler, och snart tränger de sig in från alla håll, utmanar ordet och tvingar det till förödmjukande reträtter. Tills några karismatiska bildmoralister får ny luft under vingarna, återupprättar ordets herravälde igen och skapar stora hål i mänskligheten visuella kulturarv. För ett hungrigt öga framstår 1500-talets reformerade bilder som en bedrövligt mager kost jämfört med 1400-talets virtuosa bildlekar. Konsten förföll till propaganda – till teologiska traktat och skolmässig didaktik.

*Läseboken*

Varför denna ständigt återkommande misstänksamhet mot bilderna? Varför denna djupa ambivalens inför dem? Man ser den överallt. Redan som små barn övas vi i att misstro dem. Jag var nyss fyllda fem när min mamma fick för sig att försöka lära mig att läsa. Jag låg bredvid henne i soffan i vårt vardagsrum och lyssnade på hennes röst och följde hennes finger när det rörde sig mellan bilder och bokstäver i en gammal läsebok som hon själv haft när hon gick i småskolan på 30-talet. Vi började lydigt i början av boken och arbetade oss framåt. När vi nått fram till en sida med en bild av en sjömanspojke och gått igenom de olika sj-ljuden hade jag på något magiskt sätt börjat förstå hur alltihop fungerade och fortsatte därefter mitt läsande på egen hand.

Det fanns en sorts symbolisk bildningsgång i bokens struktur. På sidorna i början var både bilder och bokstäver stora och färgglada, men allteftersom bokstävernas ljudvärden avverkades förändrades detta. Tyngdpunkten försköts långsamt och målmedvetet från stort till smått, från bilder till bokstäver, från färg till svartvitt. I slutet av boken var bilderna nästan helt försvunna och sidorna fulla




8. **BILDER I ORDETS TJÄNST.** *Läsebok för de små. Första skolåret*. Tionde upplagan 1938.

av text. För barnet som nyfiket bläddrade framåt blev det uppenbart att läsandet hade ett pris: man berövades bilderna.

Jag minns inte längre så mycket av teckningarna inuti boken, förutom den magnifika förstasidan, med en vacker åsna och en trolskt lysande måne. Med omslaget var det annorlunda. Där stoltserade en tupp på en skattkista med ena foten förnämt utsträckt framför sig och med en nyckel hängande i en tråd från en av tårna. Det var inte så svårt att se att nyckeln passade i låset till kistan och att tuppens konstlade pose uttryckte en sorts frestelse: ”Se här, denna nyckel kan bli din!”

Så formulerade den lilla omslagsbilden vår skriftförgiftade civilisations grundtanke: att konsten att läsa är nyckeln till rikedom. Samtidigt ingav tuppen mig en hel del motstridiga känslor. Trots att han inte var vänd åt betraktaren utan såg åt höger, ut ur bilden, kände jag att han ville mig något. Det var *mig* han erbjöd nyckeln. Men den lilla vassa klon som höll fram den den var ingen varm och kärleksfull människohand. Tvärtom kändes den lite oberäknelig, rentav farlig. Och ändå framhärdade tuppen där, envist pockande på min uppmärksamhet, fastfrusen i sin frestande gest. Påtagligt omtänksam, men också lite befallande. Inbjudande, men också opålitlig och lite skrämmande.

Med ett minimum av uttrycksmedel utnyttjade den lilla bilden alla de krafter som bilder kan besitta. Tuppen väckte ett slags behov i mig. En längtan. Den ville mig något. Den hade något som jag ville ha. Och jag var fast.

Vagt uppfattade jag att boken erbjöd en resa från barn till vuxen, men jag var förstås helt ovetande om att syftet med resan var att förvandla mig från barbar till civiliserad, från oupplyst till upplyst, från hedning till troende. Först många år senare har jag förstått att det bilden av tuppen utsatte mig för i själva verket var en av många interventioner som finslipats under tusentals år av skickliga domptörer som föräldrar och ledare, präster och intellektuella, experter och myndigheter.

*Djur och bilder*

Så använder vi ofta bilder av djur för att försöka förstå oss själva, få grepp om vem vi är och vad vi är med om. Som om det finns viktiga saker som vi är oförmögna att se utan deras hjälp. Genom att karikera oss hjälper de oss att tänka bättre. Det är precis som om vi blir tydligare för oss själva när vi projicerar oss i dem. Kanske skulle man kunna säga att utan de andra djuren skulle vi inte bli människor? Vi behöver dem – både som speglar och som underlägsna. Som rena varelser, befriade från all mänsklig förställning, har de en enorm kommersiell potential, men de erbjuder också det som många drömmer om: en kärlek möjlig att kontrollera.

Man skulle kunna tro att djuren som levande varelser är mer bångstyriga är andra saker som vi använder för att tänka, känna och fantisera. Men kanske är det en illusion? Det är inte bara därinne i djuret som det verkar finnas något som tittar ut genom den människomask vi tilldelat dem. Inuti de saker som vi själva tillverkat finns också ett liv som vi inte riktigt behärskar. Vi formar våra bilder för att förstå oss själva. Men när de väl tagit form får de ett eget liv.

Kanske är det därför de så envist återvänder efter alla ikonoklasmer? Vi misstror dem, men vi vill inte vara utan dem. De utlovar något. De erbjuder oss en fristad. En muromgärdad plats att vila ut på. I full uppmärksamhet. Kanske var det detta som drev Brueghel till att måla sina många blomstervaser. Hans buketter var plockade i paradiset. Och han målade även många tavlor med just paradiset som motiv, så många att han ibland kommit att kallas Paradis-Brueghel. Motiven varierade mellan Syndafallet, Adams skapelse och Utdrivandet ur Paradiset, men i några av hans över etthundra versioner av platsen spelar Adam och Eva en undanskymd roll. Här befinner sig den Heliga Skrift på reträtt. Landskapet är inte bara en torftig kuliss bakom scenens huvudaktörer, så som det är i det reformerade måleriet. Det är rentav det huvudsakliga motivet. I en av bilderna skymtar man ett par nakna figurer under ett träd, och längre bort anar man ryggen på några pyttesmå gestalter som avlägsnar sig springande, nästan

9. **PARADISISKA BILDER.** En av Jan Brueghel den äldres drygt etthundra paradisbilder, med Adam och Eva långt i bakgrunden, dels när Eva erbjuder Adam kunskapens frukt, dels själva utdrivandet. Vi är långt från skildringen i hertig Johan av Berrys tidebok (bild 1).

osynliga mitt i ett väldigt och detaljerat skogslandskap, befolkat av naturtroget återgivna djur och blommor, och man gissar att det är de båda ärkesyndarna som först plockat till sig den förbjudna frukten och sedan jagas ut ur det paradis som så vällustigt erbjuder sig åt betraktaren av målningen. Det är nästan som om det är Ordet självt som schasas ut ur bilden. Ut ur måleriet. Ut ur Guds egen trädgård.

*Bildens återkomst*

Olika tider har alltså förhållit sig på olika sätt till den ambivalenta bilden. Det går att se övergripande utvecklingsmönster här. Bilderna försvann när de kristna kulturerna ersatte de antika. De kristna patriarkerna hade inget till övers för den romerska visuella kulturen. Bildhantverket förföll i Europa. Men tusen år senare, och efter ännu en ikonoklasm under 700-talet, hade det inte bara förmått kravla sig upp ur dyn utan till och med lyckats liera sig med poesin, litteraturen och filosofin och exploderat i en så magnifik bildrenässans att vi fortfarande befinner oss i dess efterdyningar. Hur lyckades bilderna göra en så formidabel comeback?

Underligt nog verkar det ha börjat som en underjordisk verksamhet inuti skriften själv. Bokstavligen i och mellan benen på bokstäverna i missaler och antiphon. Bildviljan fick fäste i bokstavens materialitet. Dess volymer och arkitektur. Kanske började det med den ornamenterade anfangen. Man ville skapa orienteringspunkter i bokstavsflödet. På något sätt markera inledningen av en ny dag i den liturgiska kalendern, som kanske också var en händelse i jungfru Marias eller något skyddshelgons liv. Den inledande bokstaven blev till anfang och fick växa sig stor, ornamenterad och färgrik. Så småningom fick den även kropp och liv. Staplarna blev ormar och fantasivarelser. bokstäverna blev arkitektur, skulpturer, rumsskapare. Ett S kunde vara både bokstav och dubbelhövdad orm eller drake och skrivbord åt kung David. Öglefälten i bokstäver som D, P, O, R och B blev öppningar. Genom pergamentets blanka mur kunde de bokmålande munkarna nu plötsligt ge läsarna tillträde till helt andra världar, låta


10. **FÖRSTA STADIET I BILDERNAS ÅTERKOMST:** bilderna vaknar i bokstäverna och till och med bokstäverna själva börjar vilja vara bilder. Zoomorf S-initial med kung David i bön. Målad av Ingeborg-Psaltarmästaren, cirka 1205.


11. **ANDRA STADIET I BILDERNAS ÅTERKOMST:** bilderna myllrar ut i marginalerna och börjar samarbeta med varandra. Ur det frö som Set planterade i Adams mun växer nådaträdet upp och bryter sig igenom gravlocket av sten. Nertill syns Abraham i färd med att på Guds order offra Isak på altaret. En hand skjuter ner från ett litet moln och hejdar honom mitt i hugget. Ur Katarina av Cleves tidebok, cirka 1440.


12. **TREDJE STADIET I BILDERNAS ÅTERKOMST:** bilden reducerar texten till ett bihang. I Trinitasbilden i Spinola tidebok (1510–1520) är texten inskränkt till några få ord på en liten lapp, till synes fastnålad på bilden.

dem ta del av andra berättelser än dem som orden på sidan förmedlade.

Allteftersom seklerna gick blev de bildbärande anfangerna allt mer svårkontrollerade, oblyga och expansiva. Tills bilderna helt och hållet bröt sig loss och svärmade ut över sidorna, fyllde marginalerna, trängde sig in bland orden. Ingenstans var deras triumf över orden så fullständig som i 1400-talets flamländska tideböcker, där omkastningen når sin fulländing och texten till sist reducerats till ornament.

En liknande ordets reträtt ser man i senmedeltidens altartavlor och i måleriet i stort. Textremsorna fladdrar fritt fram till 1400-talet, för att sedan fasas ut samtidigt med den bakgrundsvägg av guld som de bysantinska ikonerna i tusen år hade föreskrivit måste finnas bakom madonnor och helgon. Nu blir det ordens tur att gå under jorden. Skriften blir allt hemligare, allt mer insmugen, och till sist närvarande bara i form av signaturer eller som litterär förlaga. In kommer naturtrogenheten, ljuset, skuggorna, detaljrikedomen, de illusionistiska greppen, perspektivet, bildgåtorna. Den mogna rensässansens änglar står inte längre och håller upp några stärkelsestinna språkband från en unken medeltid.

### *Mediernas kamp*

Så återföddes den europeiska bildkonsten, på nytt fri att utforska sitt eget väsen som förmedlare av andra verkligheter. När bildytan väl hade blivit en öppning i boksidans, kyrkorummets eller pannåns plana vägg, ett hål som gjorde det möjligt att kika in i ett tidigare okänt angränsande rum, en parallell rymd, uppstod det i detta rum snart samma sorts öppningar mot än djupare liggande rum, kanske en stad eller en trädgård eller ett landskap. Dessa öppningar i de illusionistiska rummen kunde sedan växa och bli till egna bildmotiv. Ur bilder föddes helt nya sorters bilder och bildgenrer, som porträtt- och landskapsmåleri.

De mest självmedvetna och sofistikerade illusionistiska strategierna utvecklades alltså i tideböckerna, vars virtuosa bildlekar lika mycket ägnade sig åt att reflektera över och kommentera det egna mediets återerövrade förmågor

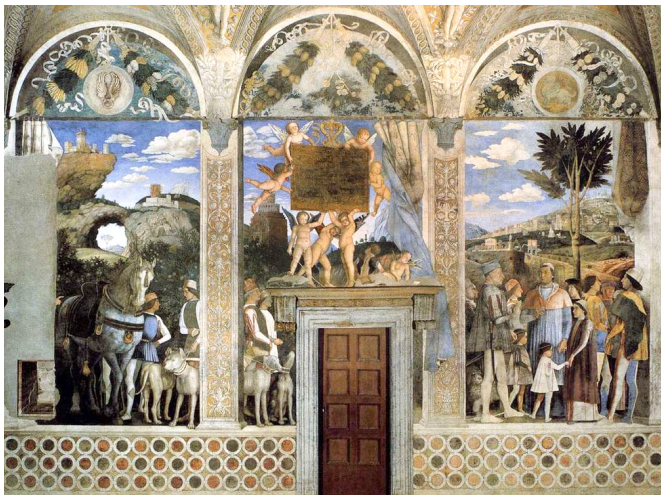
13. **DJÄRVA ILLUSIONISMER.** I en av bilderna i Spinola tidebok (1510–1520) har målaren placerat texten på en skylt med gångjärn, som om syftet vore att få betraktaren att föreställa sig att den kan svängas upp så att den inte längre är i vägen för bilden. Hela scenen är innefattad i en sorts skåpsarkitektur av trä. Ur den drömmande Jesse växer det släktträd som i toppen blommar och ger upphov till Jesus. En visualisering av det profetiska ordet om frälsaren av ”Jesse rot och stam”. Det sägs ibland att det finns fler tideböcker kvar i dag än något annat föremål från senmedeltid och renässans. De var förmodligen mycket vanliga, kanske för att många lärde sig läsa med hjälp av dem. Det är inte utan att man avundas den som fick ha en bok som denna som sin första läsebok.

som att med olika finurliga medel marginalisera texten, sätta den på plats, och emellanåt med ett skadeglatt flin rentav håna den.

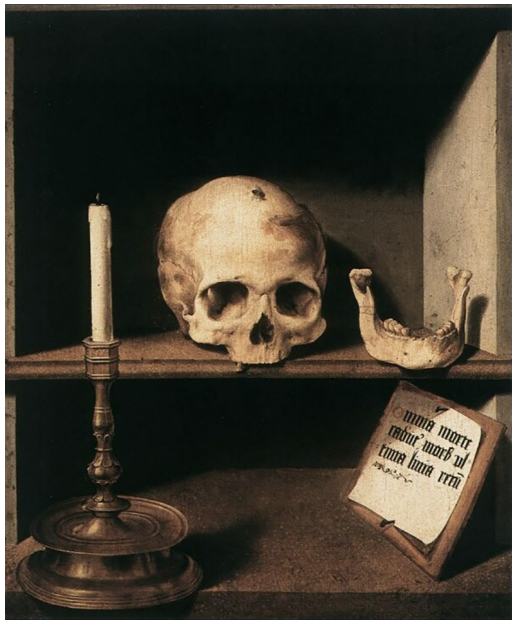
### *Den gränsöverskridande bilden*

Det är svårt att inte få uppfattningen att bildkonsten under 1400-talet närmast blir besatt av sin egen förmåga att med hjälp av detaljerad naturtrogenhet, linjeperspektiv och olika arkitektoniska markeringar lösa upp alla former av väggar. Så många av seklets bilder uttrycker nästan en sorts förälskelse i den egna gränsöverskridande förmågan, och verkar ha svårt att låta bli att mer eller mindre öppet tematisera den. Alla dessa bilder som plötsligt uppfyller konsten, måna om att vilja få oss att särskilt uppmärksamma den gräns som skiljer bildvärlden från den rymd som betraktaren befinner sig i. Själva öppningens tredimensionella gränsvärld lyfts fram: balustrader, trösklar, fönsterbräden och fönsternischer. Man ser det om och om igen, och i de mest skiftande typer av bildkonst. Alla dessa fötter som överskrider gränsen, som skjuter ut ur bildens rymd in i betraktarens. Det vimlar av dem i de illuminerade böckerna, i muralmålningarna, i kyrkorna och privatpalatsen, i altartavlorna, andaktsbilderna och porträttkonsten. Ingen bildkonstgenre tycks fri från dem, eller från något annat som hänger ut över kanten: kjortelfållar, påfågelsvansar, mattor, ljusstakar, krukor, brev, dödskallar, pallar, tallrikar.

Ibland sker det förstulet, men lika ofta demonstrativt, som i Mantegnas väggmålningar i Camera degli Sposi, bröllosrummet i Palazzo Ducale i Mantova, där bildgestalterna självsäkert skrider fram balanserande på kanten av en upphöjd scen som omger hela rummet, och mot slutet av seklet nästan som ett tvångsbeteende, som hos Carlo Crivelli i hans *Bebådelse med Sankt Emygdius* från 1486 där det firar en svåröverträffad triumf, och senare hos Barthel Bartholomäus Bruyn den äldre, i vars vanitasmålningar varje föremål helt enkelt *måste* skjuta ut över en kant. Greppet hänger sig envist kvar ända in i 1600-talet, som hos Caravaggio, där en slarvigt placerad tallrik kan kasta en oroande skugga över en bordskant, eller där en pall kan



14. **ATT BALANSERA PÅ KANTEN.** Andrea Mantegnas illusionistiska rumsarkitektur i Camera degli Sposi, bröllopsrummet i Palazzo Ducale i Mantova, 1465–74. Det är svårt att avgöra, även på plats, vad som är målat och vad som är verkligt.



15. **PREKÄRA PLACERINGAR.** Barthel Bartholomäus Bruyn den äldres *Vanitas*, föremålsmässigt återhållsam, men lika tvångsmässigt kantöverskridande som Crivellis *Bebådelse* (bild 16).


16. **UTSTUDERADE PASSAGER.** Carlo Crivellis *Bebådelse med Sankt Emygdius*, 1486.

tillåtas vara på väg att falla mellan dimensionerna, nästan på samma sätt som en ljusstråle på en bild tidigare kunde falla från himlen och befrukta en utvald jungfru.

*Kontakt*

Den som befinner sig på andra sidan bildytan går under 1400-talet allt självmedvetnare och djärvare fram till öppningen och ställer sig i den. Som för att demonstrativt förbinda konsten med verkligheten. Även när gränsen inte regelrätt överskrids markeras den ofta av en arm, en hand eller ett par fingertoppar.

Det är som om varelserna på den andra sidan av bildmembranet under 1400-talet plötsligt upptäcker öppningarna och ställer sig i kö för att gå fram till dem. Vi kan se samma sak i tusentals bilder från tiden. De står där till vår beskådan. Eller för att söka vår kontakt. Till och med Kristus ställer sig under seklet allt oftare där. Låter sina fingertoppar nudda lätta mot den tunna brädan mellan oss och honom. En bräda som som på en och samma gång är en sorts fönsterbräda och bildram.

Som om bilderna ville visa att bildens värld – kanske Jesu födelse, Marie bebådelse, eller rentav Paradiset – finns alldeles intill oss. Som om pannåerna, dukarna och bokmålningarnar ville framhäva sig själva just som membran, som gräns mellan här och där. Men en gräns som upprättar kontakt. Som gör det man ser i bilden till ett verkligt fönster i den vägg den befinner sig på.

Det är precis som om alla dessa bilder ville få oss att öka vår medvetenhet om hur vi förhåller oss till dem och få oss att börja reflektera över konstens natur och transformativa förmåga. Som om de gång på gång ville framhäva sig själva som överlägsna förändrings- och frälsningsredskap. Eller som om de ville göra oss osäkra på vad som egentligen är verkligt. Göra oss vilsna. Försätta oss i ett gränstillstånd av ovisshet där andra transformationer blir möjliga. Som om de erbjöd en väg ut ur tidens och rummets begränsningar.

”Kom in”, tycks bilden säga. ”Träd in genom porten till den hemliga trädgården. Bakom dessa skyddande murar kan ingenting hota dig.”

17. **HANDPÅLÄGGNING FRÅN ANDRA SIDAN.** Hans Memling, *Kristus välsignar*, 1481.

Så tycktes alltså bildkonsten vid övergången till 1500-talet vilja leka med tanken att den kunde erbjuda en genväg tillbaka till den lustgård som människan i tidernas begynnelse så brutalt kastats ut ur. Ett transformerande nålsöga som var och en kunde slinka ut genom för att få en försmak av himlen. Inte undra på att skriftens väktare började vakna. De nya självmedvetna bilderna började verkligen gå för långt. De måste sättas på plats, antingen elimineras eller dompteras.

Luther valde det senare och såg till att hans översättningar av Bibeln försågs med omsorgsfullt förenklade bilder. Hans epokgörande bibel från 1534 innehöll inte mindre än 123 illustrationer, var och en auktoriserad av honom själv och reducerad till textstöd, ett lydigt redskap för förståelsen av Ordet. I Luthers stränga bildregi skulle det inte få finnas något som inte nämndes i texten, inga överflödiga detaljer, ingen sublim gåtfullhet, inga löften om transformation. Bara upprepning och övertydlighet.

Kompletterande didaktiska grepp använde man i de lutherska kyrkorummen, där bilderna i fortsättningen antingen försågs med förklarande utdrag ur den heliga skrift eller helt sonika ersattes med särskilt viktiga bibelcitat. I samma förgyllda ramar som nyss omgärdat madonnor och jesusbarn upphöjdes nu skriften i form av prydliga inskriptioner av trosbekännelsen, de tio budorden eller Fader vår. Altartavlorna började bli misstänkt lika griffeltavlor. På så sätt kunde vem som helst med sina egna ögon konstatera att ordet i den reformerade kyrkan hade återtagit herraväldet och bokstavligen trängt ut bilden.

Bilderna fick alltså se sig antingen eliminerade, domesticerade eller invaderade av orden. Man kan observera den här brutala omkastningen hos en av den transformativa bildens mästare, Albrecht Dürer. Före reformationen verkar hans bilder eftersträva all den suveränitet och magiska kraft som Luther försökte undvika. Ett grafiskt blad som *Melencolia I* låter sig inte avläsas på ett enkelt sätt. Scenen den visar har ingen uppenbar förlaga i form av en helgonlegend eller ett bibelställe. Det tar en god stund bara att identifiera de olika varelserna och föremålen i bilden.


18. **ENDAST FÖR DE INVIGDA?** Dürers *Melencholia I* är detaljrik och mystisk och tycks avsiktligt skapad för att låta sig avkodas av kännare och specialister. Samtidigt är det svårt för den oinvigde att avstå från att spontant söka en förklaring som skulle kunna binda samman de disparata objekten till en meningsfull helhet, ungefär som den som drömmer inte kan låta bli att sy ihop berättelser ur det kaotiska flödet av drömbilder.

Ännu längre att göra sig en föreställning om vad de gör där, tillsammans i samma gåtfulla bildrum. Och säkert skulle man kunna ägna ett helt liv åt att försöka utröna vad bilden egentligen betyder. En enda liten bild, men ändå så detaljrik och outgrundlig att den femhundra år efter sin tillkomst genererat spaltkilometer av tolkande texter.

Sådana bilder hade Luther ingen användning för. Han var inte intresserad av bildgåtor. En bild måste klart och tydligt redovisa vad den vill. Inte försöka slå mynt av sin egen inneboende ambivalens och inbilla betraktaren att den sitter inne med livsavgörande hemligheter eller erbjuder direktkontakt med Frälsaren. Märkligt nog verkar det som om Luthers kritik av den etablerade kyrkan och av dess bilder gjorde starkt intryck på Dürer, och sent i livet ansträngde han sig för att försöka tänka i nya bildbanor. I sin sista oljemålning, lämnad som minnesgåva till stadsfäderna i Nürnberg, försökte han gestalta kärnan i den lutherska reformationen, det evangeliska Ordet och dess apostlar, men för säkerhets skull, och med Luthers bildprogram som modell, också förse bilden med en tolkningsnyckel i form av fem bibelcitat.

Dürer var noga med att förklara för stadsfäderna att det han gjort inte var någon andaktsbild. Målningen var inte ens konst, menade han, den vara bara ”ett minne”. Det hjälpte förstås föga. Eftervärlden ville inte se den så. *De fyra apostlarna* har gripit den ena generationen efter den andra av ikonofiler, och liksom *Melencolia I* har den genererat oöverskådliga textmängder. De tillfogade bibelcitaten nederst i bilden undanröjde knappast några tveksamheter. Tvärtom fick de bildens mysterierium att tätna ytterligare. Kanske för att ord, tvärtemot hur Luther tänkte sig, är lika svåra att tygla som bilder. Liksom för övrigt det mesta annat som människor skapar.

## *Utopiska bilder*

Kanske gjorde Dürersamlaren och katoliken Maximilian I av Bayern det enda raka när han hundra år senare köpte målningen, sågade av bildtexten och skickade tillbaka den till Nürnberg. Först 1922 fogades text och bild samman igen till sin ursprungliga gåtfulla helhet och visades på

19. **KRYPTISK TOLKNINGSNYCKEL.** Diptyk av Albrecht Dürer, *De fyra apostlarna*, 1526. Allra längst ner i bild, på den illusionistiska kanten till avsatsen som gestalterna befinner sig på, citerar Dürer ur Martin Luthers nya bibelöversättning från 1522: om falska lärare och profeter (Andra Petrusbrevet 2:1–3, Första Johannesbrevet 4:1–3), om självviskhet och att bära fromheten som en mask (Andra Timotheosbrevet 3:1–7) och om religiösa hycklare som ”äter änkorna ur husen och ber långa böner för syns skull” (Markusevangeliet 12:38–40).

Alte Pinakothek i München, där det än i dag kan beskådas.

För i början av 1900-talet hade det blivit självklart att det var på museum som en sådan bild hörde hemma. Inte i kyrkor och inte i stadshus, utan i särskilt utformade helgedomar ägnade åt den nationella kulturens bildkonst och dess eventuella urtida utländska föregångare. I sådana tempel fanns det inte plats för pinsamma pedantiska altartavlor från 1500-talet. Inte ens som kuriosum. Och knappast heller för de torra teologiska bildtraktat som reformationskonstnärerna producerat i sina verkstäder. Äkta konst kan så mycket mer än så. Den lyfter själen, sänder meddelanden från det osynliga, lockar med löften om sublima tillstånd. Luther hade nog blivit bekymrad. Vart hade den undervisande bilden tagit vägen i det sista seklet av det andra millenniet efter Frälsarens födelse? Fanns det inte längre någon som använde den för att vägleda människor till medvetenhet och myndighet? Hade man glömt bort konsten att tämja bilder för att med deras hjälp valla in de vilsegångna fåren på den enda sanna vägen till paradiset?

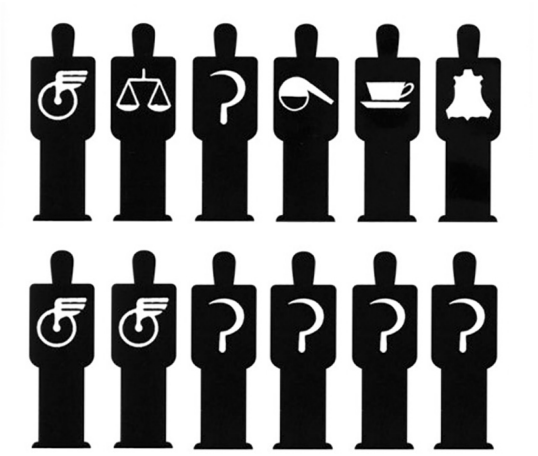
Underligt nog tycks den domesticerade bilden under 1900-talet ha hamnat i händerna på helt andra grupper av sanningssökare än de som den gamle wittenbergaren tillhörde: de samhällsomstörtande informationsgrafikerna. En av de mest radikala var förmodligen folkbildningsfilosofen Otto Neurath. Vidden av hans kompromisslöshet blir tydlig när man betraktar den bild som han ville göra till grundordet i ett universellt språk, som han kallade Isotype: ett slags esperanto i bilder i vilket man skulle kunna beskriva alla de elementära mänskliga och materiella omständigheter som en medborgare i ett modernt samhälle behövde känna till för att fullt ut kunna delta i samhällsprocessen. Neuraths folkbildande ambitioner var inte mindre än Luthers: båda ville rädda sina medmänniskor från att vandra i mörker. Båda ville hjälpa dem att bli medvetna och myndiga människor. Och båda förstod att det skulle kräva att man inrättade särskilda rumsligheter där man kunde bedriva effektiv pedagogik, och att man använde sig av bilder som alla förstod.

Isotypefiguren var en sådan bild. Betrakta den en stund. Och jämför den sedan gärna med Dürers *Melencolia I*.

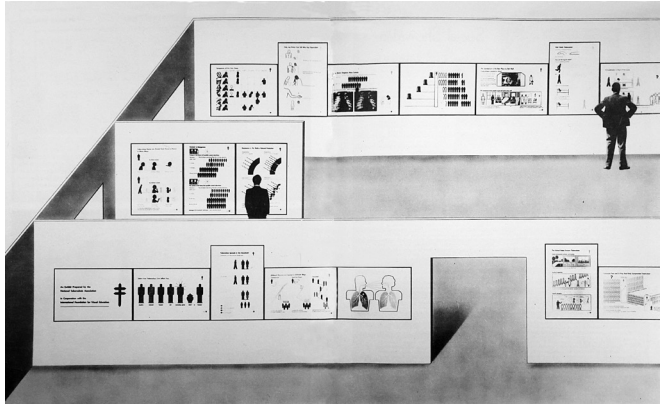
Kanske känner du dig rätt snart färdig med den ena, men dröjer dig kvar lite längre hos den andra? I så fall är du precis som jag. Det finns helt enkelt mer att titta på hos Dürer. Det händer något där i bilden, men det är oklart vad. Och det är fullt av saker där, men man förstår inte varför. Bilden tycks ropa på att vilja bli tydd, men vi vet inte riktigt var vi ska börja. Det är inte utan att Neuraths bild ter sig lite torftig vid en sådan jämförelse. Men att klaga på piktogrammets brist på uttrycksfullhet är ungefär som att klandra ett välformat A för att inte vara lika uttrycksfullt som en dikt av Tomas Tranströmer. För välformat var just vad piktogrammet var, och precis som ett funktionellt A var det modulärt och lättreproducerat, strikt utformat i syfte att massframställa meningsfulla mönster, tankar och insikter.

I sjäva verket tog Neuraths bild förmodligen betydligt längre tid att skapa än Dürers, och inbegrep en hel stab av medarbetare. Dessutom är torftigheten skenbar. Den lilla figuren bar på en mission av samhällsutopiska mått: att överbrygga klyftor och motsättningar mellan bildad och obildad, mellan klasser, språk och nationer. När vi ser den i dag tänker vi oss att den på sin höjd skulle kunna visa vägen till närmaste herrtoalett, men ingenting kunde vara mera fel. Den kom till världen för att predika och övertyga, för att skänka insikt och vinna anhängare. Den lilla bilden ville i grund och botten bara det som bilder alltid velat: förföra oss alla, såväl barn som vuxna.

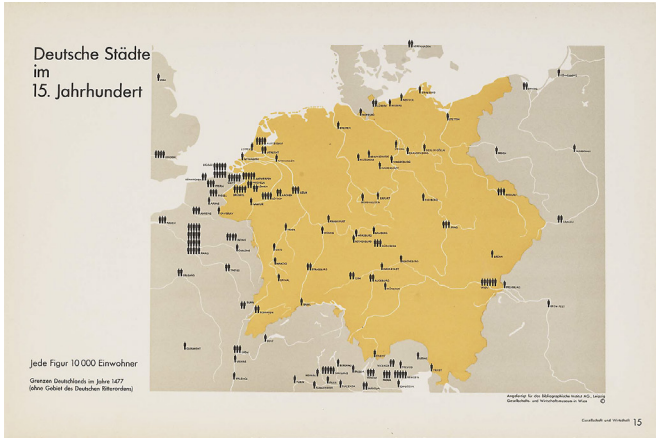
Och precis som många andra bilder krävde piktogrammet särskilda sammanhang och omständigheter för att nå sin fulla inverkan på oss. Den krävde en ny sorts tempel

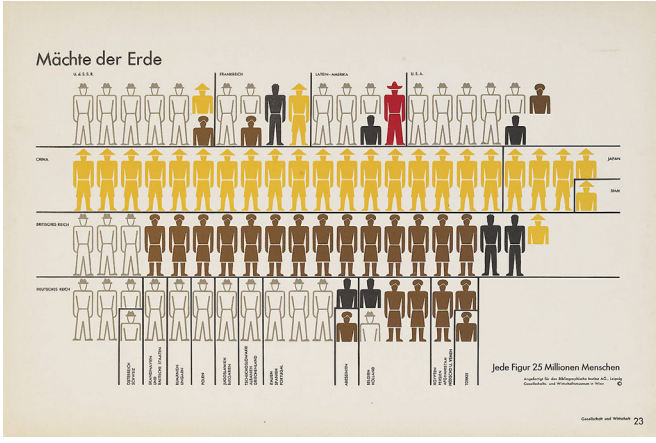


20. **ISOTYPESPRÅKETS BYGGSTENAR.** Med grafiskt tydliga och intuitivt begripliga symboler kunde grundordet i det universella bildspråket modifieras, i analogi med hur ett adjektiv modifierar ett substantiv. Genom upprepningar kunde man sedan uttrycka kvantiteter.



21. **CENTRALER FÖR VÄRLDSMEDBORGAUTBILDNING.** I Neuraths idealmuseum följde samtliga grafiska element (färger, symboler, kartor, typografi, planschformat etc.) hans stränga principer för konsekvens, modularitet, meningsfullhet och reproducerbarhet. Ett sådant museum skulle fungera som en internationell encyklopedi med uppgift att upplösa alla motsättningar mellan nationer, folkslag, klasser, vetenskapsdiscipliner och språk.


22. **STANDARDISERAT SYSTEM FÖR TÄNKANDE.** Under 1920-talet utvecklade Neurath en standardiserad grafisk metod för produktion av upplysande utställningar. Utställningen *Gesellschaft und Wirtschaft* (1930) var ett mobilt och reproducerbart museum i lådformat redo att skeppas ut över världen.

23. **UPPLYSNINGSINSTRUMENT FÖR MASSORNA.** Den utopiska informationsgrafiken har en lång och snårig historia. Buddhistiska munkar har bidragit med bland annat *Livets hjul*, en bild som i en enda grafisk struktur av koncentriska cirklar försöker visualisera Buddhas lära om obeständigheten, lidandet, karma, döden och återfödelsen. Dödsguden Yama håller allt och alla, även gudarna, i sitt både eviga och obönhörliga grepp. Bara Buddha befinner sig utanför och beskådar lugnt den onda, imaginära farsen. Bilden hängdes upp utanför klostren som undervisning för bönder och andra som inte själva kunde läsa och studera skrifterna.

som på lutherskt vis smälte samman kyrka och skola, men styrd av strikt vetenskapliga principer: en utbildningscentral. 1932 hade Neurath arbetat i tio år med att försöka åstadkomma just det, och resultatet av hans arbete fanns samlat i en liten låda som fyllde honom med tillförsikt inför framtiden. Lådan innehöll världen i koncentrat. 98 informationsgrafiska planscher i en form som till och med barn och lågutbildade kunde begripa och som gav en systematisk överblick av allt som var viktigt att känna till om människan och hennes värld. Lådan var ingenting mindre än ett hypermodernt upplysningsinstrument, en statistiskt säkerställd och vetenskapligt grundad framställning av världen i bilder, massproducerad och färdig att spridas till jordens alla länder för att packas upp och monteras på väggar i lämpligt strukturerade lokaler.

Med piktogrammet och planscherna tänkte sig Neurath att han var på väg att realisera den gamla drömmen om den fullständiga överblicken. Världen fångad i en bild, en formel eller en bok. Drömmen är densamma, men den antar många olika former. Och alltid är det någon som ska räddas. Man ser det och och om igen, i alla kulturer och i alla tider. Cranach och Luther försökte i bilder som *Lagen och nåden* på enklast möjliga sätt visualisera det enda som de ansåg var viktigt för en människa att känna till. Femhundra år senare försökte Otto Neurath göra detsamma, men i hans fall tog drömmen formen av en låda, en sällsam och portabel sammansmältning av monument och museum, atlas och utställning, begreppsordbok och encyklopedi.

## *Bildavgrunder*

En oförglömlig dag i mitten av 1960-talet stod jag utanför en kiosk i Ystad och stirrade på en av de serietidningar som hängde på rad innanför glasrutan. På omslaget till det nya numret av *Dennis* stod den lille odågan själv och tittade förvånat på en bild som han höll i sin hand. Bilden visade samma scen som man kunde se på omslaget: Dennis tittande på en bild som visade Dennis tittande på en bild som visade Dennis tittande på… En sorts svindel grep mig när jag sögs in i denna virvel av bilder inuti bilder, helt

annorlunda än de bilder jag dittills stött på i mitt högst sexåriga liv.

Samma grepp har förstås använts och varierats på olika sätt långt tillbaka i tiden, till och med i en och annan reformert altartavla, men för mig var det första gången en bild så tydligt underminerade min verklighetsuppfattning genom att öppet framhålla sin egen manipulativa kraft. Jag tror att min relation till bilder förändrades från den dagen. Och jag funderar ibland på hur människorna i norra Europa hade påverkats om de under de senaste seklerna hade lärt sig att läsa i tideböcker myllrande av illusionistiska bildlekar istället för i Luthers lilla katekes, särskilt som de allra flesta hem bara ägde en enda bok. Hur hade vårt bildmedvetande påverkats av att tidigt i livet få umgås med så sofistikerat självmedvetna bilder som de i syndafallsuppslaget i Johanna den vansinnigas tidebok, där målaren finurligt låter bilderna kommentera sin egen gränsupplösande dubbelnatur?

Det speciella med de flamländska tideböckerna är att betraktaren så öppet uppmanades att förstå bilden som ett fönster, något man ser genom snarare än på. En del av bokmålarna var uppenbarligen förälskade i denna transformativa metafor och försökte på alla vis utforska dess existentiella och filosofiska kraft i den förhållandevis fria konstform som den privata andaktsboken utgjorde. Den avväpnande lekfullheten hos bilderna i de här böckerna hjälper oss att se lite tydligare hur besatta även renässansens porträtt-, mural- och altartavlemålare faktiskt var av uppgiften att utveckla bildens potential som bro till en högre form av verklighet. Därav alla dessa naturalistiska effekter som konsten lärde sig att behärska under den här tiden: detaljrikedomen, de volymskapande skuggorna, förmågan att ge illusionen av djup och avstånd genom atmosfäriskt perspektiv, skapandet av en övertygande rationell rymd genom linjeperspektiv, naturstudierna av människokroppen i vila och rörelse, chiaroscuron, de oändliga varationerna på temat med bilden som en öppning till en parallell rymd.

Men varför allt detta slit för att odla bildens inneboende ambivalens, varför detta behov att maximera dess förmåga att trolla fram andra verkligheter, förflutna och


24. **PÅ FEL SIDA OM BILDEN.** Vänstersidan av detta strategiskt tvetydiga uppslag i Johanna den vansinnigas tidebok visar dels den välbekanta syndafallsscenen genom ett bildfönster i något slags kyrkorum, dels utdrivandet ur paradiset som ett utdrivande ur bilden, ner för trapporna, in i betraktarens egen verklighet. Uppslagets högra sida förtydligar tanken genom att i en liten medaljongspegel återge bokläsarens verkliga ansikte, dödskallens. På så sätt komprimeras den kristna berättelsen om världen till en sofistikerad bildgåta: människan har genom egen förskyllan drivits ut i en värld där döden regerar. Syndafallet är ett fall ur bilden till en värld av bilder som människan alltid är på fel sida om.

framtida, överdådiga och övernaturliga? Kanske för att man var övertygad om att bron också ledde till en högre form av människa, en tro som sedan gått i arv till generationer av målare, samlare och publik. För visst är vi på något sätt fortfarande kvar där Breughel lämnade Borromeo när han i ett brev till kardinalen bad honom ställa sig frågan ”om dessa målade blommor inte är överlägsna guld och juveler”. Innerst inne tillhör nog de flesta av oss ikonofilins trossamfund, om än med olika grader av förhoppningar om bilders transformerande makt.

Det intressanta med Otto Neurath och hans upplysningsprojekt var att han så tydligt insåg att han för att kunna skapa ett universellt bildspråk måste göra precis tvärtom. För honom gällde det att hitta metoder som reducerade bildens ambivalens till ett minimum. Helst ta kål på den helt och hållet. Det var därför han i sitt upplysningsprojekt inte hade plats för alla de naturalistiska effekter som senmedeltidens och renässansens bildskapare försökte utveckla. Luther tänkte lite i samma banor. Bilder behövde inte helt förbjudas, rätt tyglade kunde de bistå ordet, men det var nödvändigt att göra dem så grova att även den med minsta förstånd skulle begripa att bilden inte var något annat än en bild. Så eftersträvade två visionärer, var och en på sitt sätt, att få fram bilder så torftiga att ingen skulle vilseledas av dem, än mindre få för sig att tillbe dem.

Isotypefigurens öde visar hur svårt det är att lyckas med ett sådant projekt. Även om världen numera drunknar i piktogram så finns det inget enskilt dominerande piktogramspråk. Och det används inte alls så som Neurath hoppades, som byggstenar i encyklopediska utsagor. Men kanske närmar sig världen ett tillstånd där isotypespråket skulle kunna realiseras? Om Google, Facebook och Wikipedia gör gemensam sak med FN och Otto Neurath kanske Isotypes fulla potential till slut kan prövas? Vem vet. Fast när inte ens Ikea lyckas gå i land med även de enklaste bildinstruktioner finns det anledning att tvivla. Någonting i bildens natur verkar helt enkelt sätta sig på tvären.

Nej, bilderna låter sig förmodligen inte tämjas så lätt, vare sig av ikonoklaster eller didaktiker. De uppfinner ständigt sig själva på nytt. Och hela tiden föds det nya

människor som sugs in i dem och beslutar sig för att ägna sina liv åt att skapa, använda eller grubbla över dem. Utan att aldrig egentligen riktigt förstå varför.

### *Aldrig tråkigt*

Kanske är det oundvikligt att bilder förr eller senare börjar tematisera sin egen dubbelnatur, så som de flamländska tideböckerna, Dennisomslaget och många andra gjort. Och kanske är det lika oundvikligt att tider av överdriven kärlek till bilderna avlöses av misstänksamhet och hat. Det är förmodligen något vi måste försöka lära oss att leva med. De mänskliga kulturerna är alltid kluvna till sina egna skapelser, om det nu gäller guld eller juveler, pengar eller teknik, gudar eller kunskap. En kluvenhet som samtidigt verkar vara den evighetsmaskin som driver själva civilisationsprocessen.

Och finns inte samma kluvenhet inom oss själva? Är vi inte innerst inne både ikonofiler och ikonofober? Vissa bilder älskar vi. De ger oss tillträde till något djupt inom oss, något som vi inte kommer åt utan dem och som vi inte vill leva utan. Andra bilder fyller oss med avsmak och vi skulle gärna både straffa dem och rensa ut dem från verkligheten. Å ena sidan brottas vi oavbrutet med att försöka få bukt med bildernas ambivalens, å andra sidan fascineras vi och förförs av den, försöker odla den och avlocka den alla dess hemligheter. Bildernas ambivalens lämnar oss aldrig någon ro. Kanske ska vi vara tacksamma för det. Tack vare den behöver vi aldrig ha tråkigt.

# Metaphysical explanation

ANNA-SOFIA MAURIN

Dear Nils-Eric. One of the issues over which you and I have disagreed over the years (disagreeing being the best part of doing philosophy, as you know) is that concerning the being or non-being of metaphysical explanations. I tend to think that there are metaphysical explanations, whereas you tend to think that there aren't. Or, to put your view more precisely, you (together with Ingar Brinck, Göran Hermerén and Johannes Persson, in your joint paper – Why Metaphysicians Do Not Explain – published in a collection of papers very much like this one, honoring Kevin Mulligan) believe that "although we clearly explain in science as well as in everyday life, we shouldn't help ourselves to an affirmative answer to the question – do we explain in metaphysics? – at least, not without a good deal of hesitation."

## *Why you think metaphysicians do not explain*

According to you, one reason – perhaps also the *main* reason – for thinking that metaphysical explanations shouldn't count as *explanations*, is that they are different, perhaps even radically different, from what we would normally call 'explanations' (in science, but also – at least insofar as we are talking about "science-like" explanations – in everyday life). In your own words:

> …you can, of course, make the word 'explanation' stand for whatever you like. But clarity matters. Concepts are analytical tools. If we want to understand the methodological

> principles of metaphysics, we should resist Dumpty rhetoric. It is not a good idea to borrow concepts imbued with the empirical view of science and use them to analyse metaphysics. Instead let us take the methodological questions seriously and ask: What do (or should) the metaphysicians do?

Since explanation is normally understood as a notion inextricably tied up with "the empirical view of science", and metaphysical "explanation" clearly isn't, to call metaphysical explanations *explanations* is to confuse rather than to clarify. Worse, perhaps (and this is me extrapolating from and adding to what you have actually said), in the long run, to throw around the title of 'explanation' without regard for its core-meaning could end up deflating, and thereby rendering more or less useless, what is now a fruitful theoretical tool, a tool able to guide us in such matters as theory evaluation and, not least, theory choice. Clearly, that is a price too high to pay. Or so you argue.

## *Explanation – scientific, everyday, and metaphysical*

What do things we normally (in science and in everyday life) call 'explanations' have in common? At first glance it would seem that the answer is *nothing*. For surely, to explain why Mary is angry, has little, if anything in common with explaining how to bake a cake or how to ride a bike, which, in turn, has little if anything in common with explaining why water has the molecular structure it does or why the moon orbits the earth in the way it does? At second glance, we realize that the *reason* you – and others – want to reserve the title of explanation to a select few of our ordinary (and extraordinary) uses of the word, is the role explanation presumably plays as a *useful scientific concept*. In the context of giving a *theory* of explanation, therefore, the question of what makes something an explanation boils down to the question of what makes something a useful scientific tool of the kind we call 'explanation'. Let us assume, therefore that there is (at least at some suitably abstract and general level of description) a single kind or

form of explanation that is in this sense 'scientific' (and let's assume – I think in line with a more or less well established norm – that what we (rightly) call 'everyday' explanations are explanations that are basically *like* scientific explanations, only less precise). A theory of explanation, i.e., a theory which tells us which of the things we call explanations are *really* explanations, is then a theory which supposedly captures what is common to explanations of this scientifically useful kind.

What, then, is typical for a scientifically useful kind of explanation? Well, that's not entirely clear. What is clear, in brief, is that most philosophers have, so far at least, agreed that a scientific explanation paradigmatically explains *why* something happened (rather than, e.g., *how* it did), that it does so by pointing to the *cause* (or the "causal context") of the happening, and that the explanation, in so doing, increases our understanding of the happening in question, typically by allowing us to make predictions.

The most commonly accepted theory of scientific explanation (at least in modern history) is some version of Hempel's DN-model (or IS-model, in case the relevant law is statistical). At least, looking at the literature on scientific explanation, one can with some justification claim that most contemporary models of scientific explanation *depart* from this model, and try to improve it. On the Hempelian model, for the explanans to successfully explain the explanandum several conditions must be met. First, the explanandum must be a logical consequence of the explanans and the sentences constituting the explanans must be true. That is, the explanation should take the form of a sound deductive argument in which the explanandum follows as a conclusion from the premises in the explanans. Second, the explanans must contain at least one law of nature and this must be an *essential* premise in the derivation in the sense that the derivation of the explanandum would not be valid if the premise were removed. A scientific explanation, in other words, is an argument. And it is an argument which explains by telling us why the phenomenon to be explained was to be expected.

What, then, is a *metaphysical* explanation? Well, the idea is that, just as is the case with typically scientific explanations, metaphysical explanations are answers to *why*-questions. Metaphysical explanations are often taken to be the same as truthmaker-explanations (in fact, this is how you take them in your text), i.e., they are taken as something which explains *why* a certain truth is true, with reference to that in the world, the existence of which *makes* (i.e., (relevantly) *necessitates*) its truth. Truthmaker explanations are however really only one species of metaphysical explanation, the more general kind of which is normally taken to be the *grounding explanation*. Grounding explanations hold between distinct existents *x* and *y* just in case (at least) both *x* and *y* exist, and *y* exists because *x* exists, but not vice versa. Here are some examples of cases in which the relation of grounding presumably holds (and so, where the ground (metaphysically) explains the grounded), all of which have been proposed at some point by proponents of grounding:

> Socrates' ontologically fundamental constituents (e.g., a substrate *a* and all of Socrates' properties) and the complex whole which is Socrates himself (the existence of the constituents grounds – and hence explains – the existence of Socrates).
>
> Socrates' proper parts and their mereological sum (the existence of the parts grounds – and hence explains – the existence of the sum).
>
> Socrates and his singleton (the existence of Socrates grounds – and hence explains – the existence of his singleton).
>
> Socrates' natural properties and his moral properties (the existence of the natural properties grounds – and hence explains – the existence of the moral properties).
>
> The physical properties of Socrates' brain, and that brain's mental properties (the existence of the physical properties grounds – and hence explains – the existence of the mental properties).
>
> The existence of Socrates (in his current state), and the truth of e.g., the proposition <Socrates is wise> (the

existence of Socrates (in his present state) grounds – and hence explains – the truth of the proposition).

. . .

Metaphysical explanations are very different from scientific explanations, typically understood. Metaphysical explanations occur, first of all, only given that a certain relation – grounding – exists and holds between distinct existents. Metaphysical explanation is, in this sense, a kind of dependence relation. It is, however, not just that. Take Socrates' singleton and Socrates. According to the proponent of grounding-type explanations, the existence of Socrates grounds – and hence explains – the existence of Socrates' singleton, *but not vice versa*. Now, Socrates and his singleton are however mutually dependent for their existence on each other. Therefore, although grounding always involves (existential) dependence, it doesn't reduce to it. Metaphysical explanations, moreover, are clearly *non*-causal. In fact, in all of the cases listed above (with the possible exceptions of the last one), although explanandum and explanans (the relata of the grounding relation) are assumed to be distinct existents, they are clearly very intimately related. Even though non-identical, there is a sense in which they are "of the same thing". A more neutral way of putting this is perhaps in terms of constitution. Generally, it seems, in a metaphysical explanation, that which explains also *constitutes* that which it explains. Finally, the way – if any – in which a metaphysical explanation can be said to increase our understanding is considerably less clear than it is (if it is) in the case of (typical) scientific explanation. This is partly due to the fact that metaphysical explanation is in a sense *singular*, a fact that may prevents us from connecting (increased) understanding with prediction (although there is clearly *a sense* in which, e.g., given that natural properties $x$, $y$, and $z$ ground moral property $m$, we have the means necessary to predict that, given an entity with natural properties of the same kind, that entity will also exemplify $m$).

Let's summarize. First, to earn the title 'explanation', we assume, an explanation needs to be of the kind we use to answer *why*-questions, and it needs to do so in a way,

and on a subject-matter, that makes the explanation in question play a fruitful role in scientific theorizing and theory comparison. Explanations of this kind, moreover, are typically *causal* and involve an an *increase in our understanding*. However, as metaphysical explanations are typically explanations used to account for the existence of something with reference to that which makes it up, they are *non-causal*. As they are also *singular* (or, more precisely, as they do not seem to involve any *general* element (such as as a law of nature)), they do not obviously involve an increase of our understanding. This sounds like bad news for the metaphysical explanation.

## *A reason to think metaphysicians explain after all?*

Exactly what a metaphysical explanation amounts to is currently a much discussed topic in contemporary analytic metaphysics, a circumstance which unfortunately means that my characterization above does not even begin to do the many and highly sophisticated twists and turns of that debate justice. Be that as it may: for the purposes of the following discussion, all we need to agree on is that metaphysical explanation is *singular* in the sense that it picks out a relation that holds between singular existents, but does not essentially involve anything which links its holding here to its holding also in cases sufficiently similar to this one. It is *non-causal*, and it is, in one sense of that word, *constitutive* – i.e., it holds between something and *that which makes this something up*. Metaphysical explanation, that is, is the explanation you get when accounting for *in virtue of what* – or *why* – something exists, by pointing to *its* constituents.

That metaphysical explanations are very different from the typical scientific explanation is crystal clear. What, then, would constitute a reason for nevertheless calling metaphysical explanations *explanations*? Answer: if it could be shown that metaphysical explanations, although clearly very different from the typical scientific explanation, are still a kind of explanation people use in the empirical sciences, that would constitute such a reason. For, this

would mean that, besides the typical kind of scientific explanation, there is also an *a*typical kind. An explanation which answers *why*-questions about entities of relevance to the empirical sciences by offering, not the cause of their existence, but an account of what makes them up.

Whether or not there are explanations of the metaphysical (constitutive) kind also in the empirical sciences is not clear. That there isn't is, I believe, even less well established. In fact, at least *prima facie*, it is not difficult to imagine a number of different circumstances in which why a particular phenomenon exists or happens, and is the way it is, is (non-causally) explained with reference to that which makes it up. One could imagine there being explanations of this kind in physics (in terms of the fundamental particles and forces), chemistry (in terms of molecular structure), and biology (in terms of the proper parts and their organization of individual members of some species). One might also ask exactly how we ought to describe what is going on when representatives of the so-called pharmaceutical industry explain the workings of a certain drug with reference to its underlying mechanisms. Or when a certain behaviour is explained with reference to a (criteria-based) diagnosis, like e.g., ADHD. In all of these cases (and probably in many more) explanations are provided, and then scientifically systematized into entire theories (or, at least, into ever more complex explanations), and different theories are then compared and evaluated with reference to their explanatory value. Still, in none of these cases is it clear that the explanation in question is in any way causal. Rather, what we appear to have here are precisely explanation in terms of that which makes up the phenomenon about which we inquire.

If any, or all, of these explanations – on closer inspection – turn out to be anything like the metaphysical kind of explanation, this *is* a reason to think that, besides the typical kind of scientific explanation, there is also an atypical kind of scientific explanation. And if this turns out to be the case, or so one might argue, then your reason for excluding metaphysical explanations from the class of explanations would not seem to hold.

*Where do we go from here?*

If metaphysical explanations turn out to be not just a kind of explanation we find in metaphysics but also in the sciences, we have two options. Either we continue to legislate against calling them 'explanations', or we accept that they are a legitimate kind of explanation, after all. To opt for the first alternative would seem to have the unfortunate consequence of leaving us with no obvious principled reason for why some explanations are explanations, and some are not. But, going for the second alternative, which I think we ought to, is not trouble-free, of course. If, as some might think, the value of giving explanations is closely tied to the predictions they allow us to make, including metaphysical explanations among the (real) *explanations*, may render the giving of explanations (generally) useless. Or, at least, including the metaphysical kind of explanation among those legitimately so-called, might force us to rethink why explanations are valuable.

To find out which lessons, of these and other kinds, to draw we must however first make sure that we start out with a clear understanding of what characterizes (1) explanations generally, (2) (typical) scientific explanations, and (3) metaphysical explanations. A quick look at the literature soon reveals that, on all of those points, more work is needed. And then we need to investigate more carefully to see if real-life examples of scientific explanations, like the ones introduced in an admittedly all too hand-wavy and imprecise way above, are best understood as a kind of metaphysical explanation or not. A tall order. But material for much philosophical conversation – and hopefully heated philosophical quarrel (of the good kind, of course). See you in the seminar-room!

# Is preference primitive?

KEVIN MULLIGAN

Preference, according to many theories of human behaviour, is a very important phenomenon. It is therefore somewhat surprising that philosophers of mind pay so little attention to it. One question about preference concerns its *variety*. Is preference always preference for one option or state of affairs rather than another? Or is there also, as ordinary language suggests, object-preference – preferences for one person rather than another, for one country rather than another, for one value rather than another? Another question or rather group of questions concerns the *nature* of preference. Is it a mental state, disposition, act or episode, a theoretical construct, a purely behavioural phenomenon? If it is a mental state or act, is it an intellectual, affective or a conative phenomenon? If it is an affective phenomenon, does it enjoy a positive or negative "valence"? Is preference to be understood as a relation between a person's attitudes or is it a primitive phenomenon?

Unsurprisingly, answers to these questions are often not independent of one another. In what follows, I put forward some reasons for thinking that there are three distinct types of preference and contrast two views about the nature of preference, the view that preference is not itself an intentional state but a relation between intentional states and the view that preference is mentally or psychologically primitive and enjoys its own form of intentionality. The suggestions advanced in what follows are, I hope, controversial. They are certainly not defended as fully as they ought to be.

Two major *types* of preference ascription are the instances of

(1) *x* prefers to *F* rather than to *G*

and of

(2) *x* prefers that *p* rather than that *q*

To prefer to *F* rather than to *G* is to prefer one option, one course of action, to another, to prefer to travel widely rather than to read widely, to prefer to smoke rather than not to. But one may think that this preference is just to prefer *that* one travels widely rather than that one reads widely, *that* one smokes rather than that one does not smoke. Then it seems that instances of (1) are merely a special case of type (2), which might be called propositional preference. But instances of (2) range over many things other than options. They range over outcomes and many other types of states of affairs. One may, for example, prefer that society be arranged in one way rather than another. Similarly, one's preferences for certain preferences rather than others, certain emotions rather than others, are typically propositional preferences.

The term "propositional preference" (cf "propositional knowledge"), like my reference to states of affairs, may suggest that instances of (2) should be understood as relations between a subject, on the one hand, and two states of affairs or propositions, on the other hand. But there is a less baroque way of understanding instances of (2), which goes back to Prior: "prefers that . . . rather than that . . ." may be understood as a prenective or hybrid connective, which takes a name and two sentences to make a sentence. The semantic value of such a hybrid connective, on this view, is no relation but what might be called a hybrid connector, something which resembles a relation at one end only.

One reason for thinking that instances of (1) are not simply special cases of (2) may be brought out by considering a possible analogy with the structure of intentions. For Sam to want, intend (will, *wollen*) to smoke may seem at first glance to amount to nothing more than that Sam wants or intends that *he himself* smokes. But this intro-

duces into what is intended a type of reference to a subject which is not explicitly present in the intention to smoke. In one jargon, the reference to oneself is not *thematic* or explicit in intentions. In another jargon, the subject is an *unarticulated constituent* of what is intended. If this suggestion is plausible, it seems equally plausible to say that to prefer to smoke rather than to sing is not an instance of (2). But the analogy between intending and option preference is a limited one. There is a well-known argument in favour of the view that to intend (will, desire) is in fact to *intend that*. To intend to smoke is to intend to smoke sooner rather than later or the day after tomorrow etc. What do such temporal specifications qualify? Not "intend". But, as far as I can see, no such argument can be deployed to show that option preference – where this is not understood in terms of choosing or deciding – is really a type of propositional preference.

Whether or not instances of (1) are special cases of (2), ordinary language suggests that there is a distinct type of preference, object-preference. Consider the catalogue given by the Cracow poet Wislawa Szymborska in "Possibilities"

> I prefer movies.
> I prefer cats.
> I prefer the oaks along the Warta.
> I prefer Dickens to Dostoyevsky.
> I prefer myself liking people
> to myself loving mankind.
> [...]
> I prefer moralists
> who promise me nothing.
> I prefer cunning kindness to the over-trustful kind.
> I prefer the earth in civvies.
> I prefer conquered to conquering countries.
> I prefer having some reservations.
> I prefer the hell of chaos to the hell of order.
> I prefer Grimms' fairy tales to the newspapers' front pages.
> I prefer leaves without flowers to flowers without leaves.
> I prefer dogs with uncropped tails.
> I prefer light eyes, since mine are dark.
> I prefer desk drawers.

I prefer many things that I haven't mentioned here
to many things I've also left unsaid.
[...]

Most of the preferences alluded to in the full version of this poem do not range over options or states of affairs but over objects (including pluralities of objects). There seem to be many types of object-preference, of *x* preferring *y* to *z*. One may prefer Sam to Hans, Venice to Florence, claret to Burgundy, Austrian philosophy to German philosophy; liberty to social justice; the gracefulness of Giorgio's gait to Sam's clumsiness, the legitimacy of one's nation-state to the illegitimacy of the Belgian Empire, Robert Musil's Austrian irony to Thomas Mann's Teutonic kitsch, the hell of chaos to the hell of order. And so on. Object-preferences, then, seem to be three-place relations. And, as we have seen, it is not necessary to say the same of preferences that *p* rather than that *q*.

The fact that so little attention is paid to the category of object-preference in theories of preference is probably due to the suspicion that, just as preferences of type (1) seem to be special cases of type (2), so too, examples of object-preference should be seen as special cases of type (2). Preference, it may well be thought, is essentially propositional. Von Wright expresses a suspicion of this kind:

> What is it to "prefer" country *A* to country *B*? [...] Is it not to prefer to visit *A* or to live in *A* or to trade with *A*, or something similar? Generally speaking: is it not to prefer a state of affairs with regard to *A* to a corresponding state of affairs with regard to *B*?
>
> What is a person doing when he prefers apples to pears? There are many possible answers. Perhaps he likes the taste of apples better. So he prefers the taste of apples better. So, he prefers the taste of apples to the taste of pears. The state which is characteristic of a fruit is a quality or property of the fruit. Properties, like states of affairs, are proposition-like entities.
>
> But what is it to prefer apple-taste to pear-taste, or to put it more generally, one quality to another? ... In answer to the general question, one might say that to prefer one quality to another means, roughly, to prefer a state when

> the one quality is instantiated to a state when the other is ....
>
> It thus seems to be the case that preferences between states of affairs are more basic than preferences between things, in the sense that when we explicate the meaning of a preference of the second type we do it in terms of preferences of the first type. And it also seems to be the case that preferences between states of affairs are more basic than preferences between qualities of things. But I shall not maintain that this is always and necessarily so (von Wright 1983 "The Logic of Preference Reconsidered", *Philosophical Papers*, Vol. II 70)

Von Wright, then, refrains from asserting that preference is always and necessarily (what I have called) propositional preference yet thinks that it seems to be the case that preferences between states of affairs are more basic than preferences between things. But consider a preference for Sweden rather than France. Is such a preference really always to prefer to *F* in a Sweden involving way rather than to *F* in a France involving way? One possible answer to the question: "Why do you prefer to live in/visit/... Sweden rather than to live in/visit/... France?" is surely: "I prefer Sweden to France". In such cases, the preference which is motivated cannot be the preference which motivates. But one who is sceptical about the pervasiveness or fundamentality of object-preference may well concede this but go on to claim that to prefer Sweden to France must nevertheless be understood in terms of some option preferences involving the two countries. It is true that the causal genesis of a preference for Sweden over France may be activities involving both countries. But that does not rule out the possibility that an option preference be motivated by an object preference.

Scepticism about the fundamentality of object preference may also lead one to think that value preference, to prefer freedom to social justice, is just to prefer that the first value be realised or exemplified rather than that the second value be realised or exemplified. But, once again, a good answer to the question: "Why do you prefer that freedom be realised rather than equality?" is surely: "Because I prefer freedom to equality".

What is it to prefer, what is the *nature* of preference? One answer is suggested by comparative locutions such as "liking more than", "hating less than", "admiring more than". Suppose Sam is very pleased that *p* and slightly pleased that *q*. Does this not suffice for it to be the case that Sam prefers that *p* rather than that *q*? Sam's preference, we might say, is determined by the degrees of his being pleased. Sam's preference looks like an internal relation between two degrees of being pleased. Suppose Sam is pleased that *p* and displeased that *q*. Once again his preference seems to be an internal relation. But in contrast to the first case his preference is determined by the nature of his two attitudes. Preference understood in this way as an internal relation has a number of distinctive features. Sam's preferences resemble one type of doxastic property – the property someone has when she believes that *p* and believes that *q*. To believe that *p* and to believe that *q* is not to believe that *p* and *q*. A conjunction of beliefs is not any sort of belief. Similarly, one might think, the conjunction of the two attitudes in Sam, being very pleased that *p* and slightly pleased that *q*, does not determine any attitude on Sam's part at all. The conjunction determines what is often called a preference. But preference understood in this way is not any sort of mental state or act since it is a mere relation between mental states or acts and their features, an internal relation.

That this is the case is also strongly suggested by certain views about the intentionality of attitudes and other mental acts and states. On one such view, if Sam is very pleased that *p*, then he takes it to be good or valuable, in particular, pleasant or agreeable that p – he has an impression of value. Similarly, if something pleases Sam, he takes it to be pleasant. If it pleases him very much, he takes it to be very pleasant. If he admires Maria, he takes her to be admirable, perhaps courageous, or generous. If his attitude towards Jürgen is one of contempt, he takes Jürgen to be despicable. And so on. According to a development of this view, the different affective attitudes and their axiological correlates are related in the following way: attitudes are correct iff their objects exemplify certain value properties. Then to be pleased by something is correct iff it is pleasant;

indignation that *p* is correct iff it is unjust that *p*; shame about some past deed is correct iff the deed was shameful. And so on.

If that is the case, then it is plausible to say that a preference that *p* rather than that *q* is correct iff it is better that *p* than that *q*. But from the fact that Sam is very pleased that *p* and slightly pleased that *q* it does not follow that Sam has any impression that it is better that *p* than that *q*. Indeed not only might Sam lack the concept of betterness he might lack any acquaintance with comparative value. Thus if we say that Sam's two degrees of being pleased determine a preference, we should not say that this preference is any sort of affective mental state which enjoys intentionality.

The view that internal relations between attitudes and their degrees suffice for preference fits some cases better than others. Consider a world in which the only affective phenomena are degrees of being pleased and being displeased. In such a world a person's preferences are easily determined. (But even in this world we may wonder whether the attitudes which determine a subject's preferences have to be simultaneous). This world is the real world according to one philosophy of emotions. According to a very different philosophy, positive and negative emotions come in qualitatively different kinds; admiring, approving and adoring, say, are qualitatively different. Suppose *x admires z* enormously at *t* and *adores y* a little at *t*. It is by no means obvious that these two facts determine a preference. If one thinks that emotions differ not only in degree and kind but may also be more or less deep, then it appears that emotions determine preferences only in certain very simple cases.

Is there a mentalist alternative to the view that preferences are internal relations between attitudes? Such an alternative will presumably take seriously such phenomena as impressions of betterness and, in particular, the intentionality of such impressions.

One such impression is that one person (thing, animal, country) is better, more beautiful, useful, elegant, healthy, … than the second. Here the axiological relation is an external relation. A related type of impression is that one

value is more important, a higher value, than another value. Nietzsche, for example, had the distinct impression that the value of life is more important than the value of knowledge or truth. Here the axiological relation is an internal relation. But it is an internal relation which should not be confused with the internal relations between the degrees of value (positive, negative, indifference) of things and persons, which, on one plausible view, are part of the make-up of contingent relations of value between objects. (Compare the difference between the external relation of being more or less expensive than and internal relations between prices, and the difference between the external relation of similarity between things and internal relations of distance between qualities). The different axiological relations between objects correspond to similar relations between options and states of affairs.

What, now, are impressions of betterness and importance? Such impressions may occur without explicit comparisons or on the basis of such comparisons. What is an impression of betterness? One answer is that such an impression is a judgment, in particular a judgment to the effect that one thing is better than another. Similarly, it has often been argued that emotions are just evaluative judgements. Suppose we are convinced by the arguments against the view that to emote is to judge, many of which resemble the arguments against the view that to see is to judge. Such arguments strongly suggest that impressions of betterness need not be judgmental either. What might an impression of betterness or importance be if it is not a judgment or belief ?

Perhaps an impression of betterness or importance is just a preferring. Preferring one thing to another is correct only if it is better than the other thing. The formal object of such preferring is betterness. Preferring one value to another is correct only if the first value is higher in value or more important than the second value. The formal object of such preferring is value height. Similarly, it is often thought, as we have noted, that different monadic values figure in the correctness conditions for different types of emotion (indignation and injustice, shame and shamefulness). This last claim is often combined with the view that

only emotions can reveal or disclose value properties. That seems to me to be wrong. Emotions are typically motivated and triggered by impressions of value which precede them. It is not inconsistent with this claim to think that only preferring can reveal or disclose betterness. For preferrings are not emotions. They are affective phenomena – one's heart turns in one direction rather than another – but they are not emotions. Emotions are attitudes but preferrings are not attitudes. Preferrings have no polar opposite. In this respect, they resemble judgings rather than belief. For judgings, if Bolzano and Frege are to be believed, have no opposites, although belief is opposed to disbelief, and certainty to uncertainty. Preferrings have no "valence", they are neither *pro* nor *contra* anything. In this respect, they resemble surprise. A preferring is an episode, unlike a preference. The relation between preferrings and preferences resembles the relation between judgings and beliefs. A judging typically marks the beginning of a belief. Similarly, episodic preferrings may mark the beginnings of the states and dispositions we call preferences.

The suggestion that preferrings and the preferences to which they give rise are the best candidate for the rôle of impressions of betterness and of importance has two interesting features. First, it complements the popular view that emotions or other affective phenomena (for example, *Wertfühlen*, the phenomenon of being struck by value) reveal or disclose monadic value. Emotions or impressions of monadic value and preferrings, including value-preferrings, have as their objective counterparts the full range of axiological objects, properties, relations and connectors: positive and negative value, beauty and ugliness, the relation of betterness, the relation of being more elegant than, value-height, the state of affairs that it is worse or more shameful or more unjust that $p$ than that $q$, and the state of affairs that it is worse to $F$ than to $G$. Secondly, the suggestion immediately provides an answer to the question about the origin or source of the concepts of betterness and importance. These concepts, the answer goes, have their origins in preferrings and in their "intentional objects". An alternative view of the origin of the concept of betterness is that this concept depends on a grasp of the

concept of monadic value and on the concept of more or less. But it is not obvious what a parallel account of the origin of the concept of importance or value-height would look like.

What is the relation between preferring, understood as a fully intentional episode, and other affective phenomena such as emotions or being struck by value? If, as is sometimes claimed, betterness is more fundamental than monadic value, preferring might be independent of all other affective phenomena. Another possibility is that preferring presupposes other kinds of affective phenomena. The formation of a preference for one thing rather than another presupposes some grasp of the value-properties of the two things. As we have noted, such a grasp may be taken to be disclosure by emotion or some other type of value impression. This grasp may also be purely conceptual, as when we come to prefer one thing to another on the basis of knowledge by description. But it may also be wholly intuitive as when Giorgio, on the basis of a rapid examination of two new handbags from Milan, plumps for the one rather than the other. And, of course, many different combinations of conceptual information and impressions may provide the starting point for preferrings.

The two accounts I have sketched of the nature of preference are very different. On the first account, preferences are an ontological – in particular, a psychological or mental – free lunch; they supervene on or are determined by or are constituted by a person's emotions and sentiments and their features. On the second account, preferences are brought into being by preferrings, understood as episodic impressions of betterness or importance. Are the two accounts really rival accounts? Why not think that there are preferences of both types? The existence of preferences which display intentionality and of preferences which do not seems to me to be incompatible with the idea that preference has an essence or nature. It also seems to me that the strongest part of the case for wholly intentional preferrings is the part dealing with impressions of importance or value-height. For in such cases differences of degree between monadic value properties can play no role. But here too the friend of the view that preferences are

determined by attitudes and their features has an alternative account available. He may say that a preference for one value over another is determined by the relative *depths* of a person's attitudes and sentiments. Thus the preference of an anti-Nietzschean might be determined by a deep attachment to, or reverence for, the value of knowledge and indifference to, or a superficial aesthetic appreciation of, the value of health. But, as far as I can tell, the relation between value-height, value-conflict, preference and action is still much under-explored. Some aspects of this relation, nicely formulated by Bernard Williams, make it clear why the relation is still so little understood:

> Very many of our [one-party, one-person value-] conflicts … are at a level where interpretation in action is less determinate or immediate. Values such as liberty, equality, and expressions of justice other than equality, can certainly conflict as ideals or objectives, though their connection with immediately presented courses of action may often be problematical, while, in the other direction, a choice between presented courses of action may in some cases be only indeterminately guided or shaped by appeal to these values. - Still further from particular choices of action or policy are evaluations of admirable human characteristics or virtues such as courage, gentleness, honesty, independence of spirit and so forth (Williams, B. 1981 "Conflicts of Value", *Moral Luck*, 75–76)

Something like the account I have sketched of preferring as a wholly intentional phenomenon has, as far as I know, only ever been endorsed by one group of modern philosophers – by Brentano and some of his Austrian and German heirs. In his attempt to resurrect Aristotle's account of preference and to put it at the heart of the philosophy of mind and value – for another resurrection, see the very rich paper by our birthday boy, Sahlin, N-E. 1993 "Worthy of Choice", *Theoria*, LIX, 178–191) – Brentano employs the unusual concept of a preferring (*ein Vorziehen*) and describes preferring as "a relating liking or loving" (*ein beziehentliches Lieben*). Something like the view of preference as an internal relation between attitudes was sometimes called "analytic preferring" in the Brentano tradition,

and something like what I have called value preference was there called "synthetic preferring" (Hermann Schwarz).

One of the ironies in the history of the theory of preference is that it seems to have been Brentano's Prague student, Oskar Kraus, who persuaded the early Austrian economists, in particular, Böhm-Bawerk, to introduce the concept of preference into their accounts of economic behaviour and marginal utility. But Kraus did not manage to persuade the economists to employ Brentano's account of preference. It was therefore only a matter of time before preference came to be seen as something which is no mental state but is wholly determined either by attitudes and mental states or by behavioural dispositions.

# How does your garden grow?

JOHN D. NORTON

*For Nils-Eric Sahlin on his 60th birthday*

For a number of years, I have run a center in philosophy of science. Our stated mission is to foster the emergence of the best new work in philosophy of science. We want our center to be a place to which philosophers of science come to do the best work of their lives. The idea, in the abstract, is that we have an environment that fosters creativity. The ambition is easily stated, yet on my first day in my new office, I realized that I had little idea of how to bring it about concretely.

Over the following years, by trial and error, I found some things that work and some that do not. It was a lonely pursuit. Our community does not take creativity and creative environments as an explicit topic of investigation. Attempts to open discussion of the topic most likely draw awkward responses and incomprehension.

All this changed the day Nils-Eric stepped into the Center in August 2011 to begin his year's visit as a senior fellow, the Wagner Risk Fellow. He immediately commented on the creative environment and I soon found that he did not just talk about creative environments. He also *wrote* about them and how to achieve them.

Lest this seem unimportant, let me reflect for a moment on the deplorable culture prevalent in philosophy of science. It is adversarial and combative, censoring and repressive.

Graduate students working towards a doctorate in philosophy of science are focused on multiple menaces. There

is the dissertation committee, whose whim controls their fate. While they seek to divine those whims, they also worry endlessly about generic appointment committees who will decide whether they will enter the profession with a job or fade away. How much do I need on my CV to be competitive? Will this writing sample be dismissed as unreadable because it is too dense? Or lightweight because it is too easily read? Will this job talk impress the experts and inform the rest? Or will it be a long-remembered moment of shame and confusion?

The resulting sense of powerlessness is just the beginning. New doctorates and junior professors need refereed publications, for the next hurdle is the tenure decision. They rapidly learn the truth about the refereeing system. For every referee who does a conscientious job, there's another who reads superficially and dismisses in haste, sometimes with callous cruelty. Worse, the system discourages real creativity. For no new idea emerges free of tension with the prior, often dull-witted literature. Who will referee the new idea? It is just the dull-witted authors of this prior literature. Then the anonymity of the referee's report can hide self-serving censorship. All this encourages uncreative writing that reaffirms the referees' published ideas and merely adds just enough novelty to provide cover for a recommendation to publish.

A culture of confrontation now permeates the rooms in which we give our talks. There will be occasions in which a speaker will be directly contradicting the research of someone listening. Then it is quite appropriate for a debate to open. When this happens, we have developed mores that hide the acrimony behind forced politeness. The question starts with praise of some minor point and then leads up to an apparently innocent "but I don't quite understand what you mean when you say…" Everyone understands it to be a challenge and a blank statement of disagreement.

The sad thing is that this combative interaction, masked by false politeness, often becomes the default mode of an audience, even when there is no real disagreement at hand. I sometimes hear of speakers being "put through their paces," as if they are racehorses who have taken too lazy a

turn round the course and must now be whipped to make them work and sweat. This may be excused as an effort to help speakers develop their ideas. It is no such thing. It is merely an ego-driven exercise in cruelty. No speaker should have to suffer it.

With this as our culture, when new fellows meet for the first time, it is no wonder that they are guarded and defensive. What can I do to relax them? What can I do to make the environment creative?

Nils-Eric has the recipe. It is in his "Creating Creative Environments" (in *Trust and Confidence in Scientific Research*, eds. G. Hermerén et al., Royal Swedish Academy of Letters, History and Antiquities, Stockholm, 2013. Download at http://www.nilsericsahlin.se/) and other writings referenced there. Nils-Eric lists nine ingredients that make an environment creative. I repeat his list here:

1. Generosity
2. A sense of community
3. Qualifications
4. Diversity
5. Trust and tolerance
6. Equality
7. Curiosity
8. Freedom of spirit
9. Small scale

One can already see the wisdom of his recipe merely by reading these headings. They do need some elaboration and for that I refer you to Nils-Eric's writings directly. Rather than repeat what he has already written, let me add a remark. The ingredients are of two types. One has its origins external to the environment; the other comes from within the environment.

That is, 3. Qualifications, 4. Diversity and 9. Small scale, cannot be created once a group of scholars has been assembled. That they are suitably qualified, represent many perspectives and are of the right number to engage in fertile interactions, must be brought about by whoever assembles the group. In our Center, that is the job of the Director and Center Officers when they conceive the

group activity and recruit its members. Because these ingredients arise, in effect, through administrative fiat, they are easier to put in place.

The remaining ingredients come from within the environment. Generosity, a sense of community, and so on, are all part of the culture of the group. They cannot be dictated or imposed. They can only be encouraged and will arise if the individuals of the group decide to commit to them. It requires their active participation and their consent.

My experience is that this culture can be encouraged. I have found both direct and indirect ways to do it. Directly, at our initial meeting, I will criticize the present culture in philosophy of science, much as I did above. I will then issue a blanket prohibition on gratuitous criticism. I have found a helpful slogan to be "no problems without solutions." That is, if you must criticize a fellow's work, you must also offer a solution to the problem. You might imagine that this speech is futile and even heavy-handed. Perhaps it is the latter, but it is not the former. It has been universally greeted with smiles and visible expressions of relief.

The indirect ways of establishing a communal culture are obvious, but still require effort. We must meet and do so often. We must engage in the activities of a family. We eat together, and often. To know someone across the dinner table is to know them as family. Then there are rituals. Religions have long recognized that we find rituals appealing and binding. Without going into tedious details, I will mention "umbrellas" and the installation in the wall of fame as rituals that those who have visited will instantly recall. They are consciously contrived as shared experiences that initiate fellows into the community.

There are two additions I would like to make to Nils-Eric's list.

10. Every garden needs a gardener.

If there were only flowers in the world, we would never need to tend our garden. However there are many weeds in the world. Everyone carries the weeds of the combative

culture of philosophy of science into the room with them, whether wittingly or not. Our garden needs a gardener to make sure that it is flowers and not weeds that are seeded and sprout. No matter how pretty the garden we seed in our community, it is still surrounded by the weeds of the larger community. It is easy for these weeds to spring up. It is easy to lapse into thoughtless criticism. Helping a struggling scholar with a weak idea is vastly harder than administering the obvious coup de grace. A well-motivated community will do some weeding on its own. However some slide is inevitable. It is important that an organizer be vigilant and reaffirm Nils-Eric's ingredients when this happens.

11. We are all different.

We have to allow that there will be many different personalities in the group. Some will be naturally gregarious and community-minded. They form the social core of the community. They naturally connect, know what everyone is doing and may even spontaneously organize social activities. They may well also be the collaborators who love to work together. Others, however, will be retiring. They prefer to work in isolation. They do not want to co-author papers. This preference must be respected. It is quite compatible with these scholars forming an integral part of the community. They will listen more than they speak, but they will listen.

There will be extreme cases of community members who rarely speak. I had initially regarded these as some sort of failure. I now see them otherwise. While they may listen and rarely speak, I have repeatedly been surprised to find just how much they value the experience. When the end of term comes and we must decide if this is our last meeting, they may be the ones to urge one more meeting.

⋆

It is a curious fact that something as important as the ways of building a creative community receives so little attention. What are we to make of it? It is, I believe, part of a

larger pattern in philosophy of science. We enter the discipline because we want to think, write and talk philosophy of science. As students, we are given intensive instruction in the content of philosophy of science and in the writing of papers. We are well-prepared for this part of our professional lives. However there is so much more that we will do. We will give talks. We will teach classes. We will interview and hire colleagues. We may even take our turns as administrators, such as departmental chairs. How do we know how to do these things well? There is an expectation that we do, but no structure to ensure it. We just have to pick it up on the way through. Or at least some of us will. As for the rest, we all know in our community of the terrible speaker, the mediocre teacher and disastrous administrator.

# The (misconceived) distinction between internal and external validity

JOHANNES PERSSON
ANNIKA WALLIN

## *1. Two common (but misconceived) claims about internal validity: the priority and trade-off claim*

Researchers often aim to make correct inferences both about that which is actually studied (internal validity) and about what the results generalize to (external validity). The language of internal and external validity is not used by everyone, but many of us would agree that *intuitively* the distinction makes a lot of sense.

Two claims are commonly made with respect to internal and external validity. The first is that internal validity is prior to external validity since there is nothing to generalize if the findings obtained in, for instance, the experimental setting do not hold. The first claim is explicit in many writings. See for instance Francisco Guala's influential book *The methodology of experimental economics* (2005). And it is often implicitly relied on. The second claim is that researchers have to make a trade-off between internal and external validity. When one is increased, the other will decrease. The second claim was made already from the start by D.T Campbell in his classic *Factors relevant to the validity of experiments in social settings* (e.g., Campbell 1957, 297).

There is a certain tension between the first and the second claim. It has been argued before that it might be difficult to combine them. We intend to make the stronger point that both claims are misconstrued. Our hypothesis is that the relationship between internal and external

validity has to be re-conceptualized, and we will briefly indicate how.

## 2. *Some remarks about the origin of the divide between external and internal validity*

Donald T. Campbell introduced the concepts internal and external validity in the 1950s. In this text we rely on his 1957 classic (already mentioned in the introduction) as the primary source to his conceptual pair:

> First, and as a basic minimum, is what can be called internal validity: did in fact the experimental stimulus make some significant difference in this specific instance? The second criterion is that of external validity, representativeness, or generalizability: to what populations, settings, and variables can this effect be generalized? (Campbell, 1957, 297)

The original article discussed research related to personality and personality change, but the conceptual pair of external and internal validity was soon extended to educational and social research. Since then it has spread to many more disciplines. Without a doubt the concepts capture – roughly, at least – two features of research that scientists are aware of in their daily practice. Researchers aim to make correct inferences both about that which is actually studied (internal validity), for instance in an experiment, and about what the results 'generalize to' (external validity). Whether or not the language of internal and external validity is used in their disciplines, researchers often experience the difference and sometimes the tension between these two kinds of inference. For instance, Nancy Cartwright in her *Hunting causes and using them* (2007, 220) calls the trade off between the two kinds of validity "a well–known methodological truism".

It is interesting to note that there in Campbell (1957) is no explicit mentioning of causal inference. On the other hand the language of effects is used rather extensively – as, for instance, in the above introduction of internal and external validity. What is salient already from the beginning is a strong link between the internal/external validity

distinction and the process of finding hypotheses among which a choice can be made:

> The optimal design is, of course, one having both internal and external validity. Insofar as such settings are available, they should be exploited, without embarrassment from the apparent opportunistic warping of the content of studies by the availability of laboratory techniques. In this sense, a science is as opportunistic as a bacteria culture and grows only where growth is possible. One basic necessity for such growth is the machinery for selecting among alternative hypotheses, no matter how limited those hypotheses may have to be. (Campbell, 1957, 310)

The causal vocabulary in Campbell's writings becomes more pronounced in his later production. At the same time, Campbell weakened his claims concerning the connection between local and general causal claims. There is a clear difference between Campbell 1957 and his *Relabeling internal and external validity for applied social sciences* from 1986, for instance. Partly, we think, this was because of his growing interest in applied sciences. Applied scientists also need internal validity, but they can normally not analyse causation with precision. There is a certain vagueness in the context of application. It is normally impossible to say with certainty which components of an intervention are causally relevant. This has implications for the internal/external validity distinction. At any rate this appears to be the received wisdom today, and it is reproduced in influential textbooks – such as in the *Experimental and quasiexperimental designs for generalized causal inference* written by W.R. Shadish, T. D. Cook and Campbell himself (2002).

### *3. On the priority claim: temporal and epistemic aspects*

In both introductory and more advanced methodological textbooks, it is often claimed that internal validity is both temporally and epistemically prior to external validity. An example is Francisco Guala's paper *Experimental localism and external validity*:

> Problems of internal validity are chronologically and epistemically antecedent to problems of external validity: it does not make much sense to ask whether a result is valid outside the experimental circumstances unless we are confident that it does therein (2003, 1198)

The claim about temporal priority is that we first make inferences about the local environment under study before making inferences about the surrounding world. The claim about epistemic priority is that we come to know the local environment before we come to know the surrounding world. Maria Jimenez-Buedo and Luis Miller (2010) have recently collected a number of similar claims from the literature. Two examples are: "internal validity is a necessary but not sufficient condition for external validity" (found in *The challenge of representativeness design in psychology and economics* by Hogarth 2005); and "if there are doubts or questions about whether a relationship is real or spurious, then whether or not the finding applies to other settings is irrelevant" (found in *Reliability in experimental sociology* by Thye 2000).

The rising interest in experiments and methodological issues in sciences where experimentation has not been extensively used before has pushed the internal and external validity distinction into focus, although comparatively little – indeed, surprisingly little – has been written about the topic within philosophy of science. Recently, it has mostly been addressed in the philosophy of economics, due to the rising importance of, and philosophical interest in, experimental economics.

## 4. *The curse of context*

The discussion within philosophy of economics and philosophy of natural sciences interconnect. For instance, it is claimed by Jones (2011) in *External validity and libraries of phenomena: a critique of Guala's methodology of experimental economics* that Guala is strongly influenced by Ian Hacking's characterization of laboratory sciences: "those whose claims to truth answer primarily to work done in the laboratory" (Hacking 1992). This influence, Jones argues,

leads Guala to overemphasize the difficulties of bridging *the gap* between internal and external validity. Guala makes a rather strict divide between testing for robustness (according to him this is an acceptable laboratory procedure) and testing for external validity (which he claims is impossible, due to the fact that experimenters cannot exactly reproduce the real system in the laboratory). This analogy between the natural and the social sciences is, however, easily drawn too far.

Many naturalistically inclined methodologists and researchers want to point out that there is an essential difference between the natural and social world with regard to the way the study objects are affected by different contexts. In fact, and this is our argument, this interplay is one of the things that threaten the priority claim of internal validity.

Social scientists worry that participants bring their experience of the world outside the laboratory with them into the experimental setting, and this may fundamentally change the way that the "target system" and the laboratory "reconstruction" of this system relate. What the researchers find to be internally valid results might be strongly dependent on them being externally valid, in a loose sense. We know them to hold outside the laboratory, and that is why we discover them in the laboratory. Furthermore, applied researchers within this field do not remain with the laboratory setting, something that further complicates the internal/external validity distinction. For instance, Baruch Fischhoff (1996), in the wonderfully titled "The real world: What good is it?", published in *Organizational Behavior and Human Decision Processes*, has argued that applied psychology can change the way that experimental psychology is conducted by allowing researchers to better understand the nature of the laboratory tasks. In particular, a little applied psychology may open researchers' eyes to "the curse of context" (that participants bring their own understanding to the minimalist problems set before them in the laboratory) and the "curse of cleverness" (devising complex experimental tasks with the assumption that participants immediately will understand their structure). The curse of context

clearly threatens the internal/external validity distinction by putting into question whether we can isolate that which we observe from the context. This might not be Fischhoff's worry, but it applies to the problem at hand. Fischhoff's mission is another. The standardization of stimuli in controlled laboratory settings turn participants into "battery raised hens", Fischhoff claims, with the hope of being able to produce predictable changes in output, whereas applied psychology studies free range poultry instead. Fischhoff's hope is that the combination of the two will lead to a better understanding of cognitive processes.

### *5. EBM and Vetenskap och beprövad erfarenhet (Science and proven experience)*

Interestingly some of the discussions within the philosophy of natural science and economics have carried over to the more applied field of evaluating the strengths and weaknesses of evidence-based medicine (EBM) and health care. Of particular interest in this connection might be to study concepts such as (the distinctly Swedish notion) "vetenskap och beprövad erfarenhet". Nils-Eric Sahlin recently acquired a substantial amount of money from Bank of Sweden Tercentenary Foundation, and we very much hope that that programme will shed light on the distinction between internal and external validity as well. We have started to develop some such ideas in *Vetenskapsteori för sanningssökare* (Fri Tanke 2013).

Randomised controlled trials (RCT) are often seen as the privileged route to causal inference in EBM. RCTs are important in this context since they enforce both the idea that internal validity is prior to external validity and that there is a trade-off between the two types of validity. However, we should perhaps take care to distinguish causal inference from inferences involving the elimination of alternative explanations. Hence an implication to be explored emerges from a position where it is accepted that it is not a coincidence that A and B occur together and where it remains an open question if A and B are causally related. This possibility leaves open that internal validity (A causes B in the trial) depends on the external validity of the claim

(A causes B outside the laboratory). It is interesting in this context that later Campbell proposed the abandonment of the concept of internal validity and suggested 'local molar validity' (i.e. inference to a complex package of potential difference-makers) in its stead.

### *6. Artefacts and internal validity*

The importance of how participants adapt to, and utilize, contingent features of their everyday life and bring this with them to the laboratory setting has been discussed within cognitive psychology even prior to Campbell's notion of external validity. It is often traced back to Egon Brunswik's perception research which challenged the Gestalt psychologists' focus on perceptual illusions by demonstrating a surprising degree of perceptual accuracy under natural conditions. Brunswik's insistence on performance in the natural world presupposes external validity, and has given rise to the probabilistic view on judgment and decision making that we will explore more closely below. One of his main interests was how well factors imperfectly related to a criterion to be predicted function in real life. He is, for instance, well known for research in which he tried to determine to what extent retinal size could be used to predict the actual size of an object. In principle, retinal size is not a good cue for actual size, since both objects' size and their distance to participants can vary. In practice, however, objects tend to be of certain sizes and be looked at, at certain distances. Such contingent relations can, and to some extent do, make retinal size a good cue for actual size in natural environments. Brunswik is, however, not only historically important. His emphasis on representative sampling is also a tool for identifying the situations or environments in which decision making is supposed to succeed. In his *Distal focussing of perception: Size constancy in a representative sample of situations* (Psychological Monographs, 1944), Brunswik attempted to randomly sample instances in which participants spontaneously looked at objects in their everyday life, and measure the correlation between retinal size and object size in these particular situations. The environment in which the pre-

dictive potential of retinal size is measured is thus determined through *representative sampling*.

Representative sampling is a key phenomenon in the internal/external validity debate since it emphasizes that good experimental data only can be found if the experiment is – in important respects – similar to the everyday surroundings of participants. Within, in particular, judgment and decision making research, the ideas of representative sampling have resurfaced through a relatively recent debate regarding the validity of a number of experimental findings allegedly demonstrating the inaccuracies of human judgment. The key role of external validity here is thus not to guarantee the generalizability of experimental findings (the role still exists though). Rather, the potential generalizability of the findings is what guarantees that the experimental results are not merely artefacts. This might happen both in obvious and more oblique ways.

The most obvious example is that researchers may, in the experimental task, use (or interpret) words in a way that is unfamiliar to participants, or at least different, from how participants use them. For instance, when participants are asked to state their probabilistic beliefs, 50% (.5, or similar) has an elevated frequency, presumably because phrases such as "fifty-fifty" are taken to represent uncertainty rather than a particular probability, as was established in the paper *Fifty-Fifty = 50%?* (Fischhoff & Bruine de Bruin, 1999). Sometimes differences in terminology have been argued to be the true cause of well-known experimental effects. With respect to the conjunction fallacy (related to the famous Linda-problem), it has repeatedly been argued that the fallacy is due to participants' (mis) understanding of "probability" when participants are given the task to rank statements "by their probability", or of the operator "and" when participants then rate the critical statement "Linda is a bank teller and an active feminist" to be more probable than "Linda is a bank teller".

There are, of course, many more examples, but the main point is that potential experimental artefacts such as these demonstrate that participants bring their knowledge of the surrounding world into the laboratory. In so far as the

experiment, or the experimental stimuli, in some important respect misrepresents participants' experiences, it is likely that the behaviours observed in the laboratory are mere artefacts. In these cases, both internal and external validity are compromised. Truly internally valid results require that we see clearly, i.e. that what we see in the local environment is not in fact an artefact of something else. And to be able to identify the experimental artefacts, we need to be able to see what participants see – a skill that can be trained through applied research, according to the argument of Fischhoff above.

### *7. Two problems*

From the above one can argue that the claim that internal validity is prior to external validity is too simplistic by pointing to two epistemologically problematic aspects: experimental artefacts and the implication of causal relations. Each demonstrates how important external validity is to the internal validity of the experimental result.

For instance, if the aim of an experiment in psychology is to understand the functioning of different psychological mechanisms (in the form of stimulus-response relations), then the quality of this finding is just as dependent on whether the psychological mechanism has been properly activated as it is on whether the results can be replicated. This is not only a question about how the result will generalize to other settings (external validity) – it is a question about whether a proper result has at all been generated (internal validity). Thus, for psychological mechanisms that can be assumed to have an adaptive character, external validity (or certain aspects of it) appears to be prior to internal validity: It is more important that an experiment measures what it aims to measure than that the result is internally valid. Egon Brunswik puts it neatly: "psychology has forgotten that it is a science of organism-environment relationships, and has become a science of the organism" (Brunswik, 1957, 6).

# Becoming our selves

JOHANNA SEIBT

We, human beings, are entities with the characteristic feature of self-awareness: we are present to ourselves. More cautiously speaking, we are entities who experience ourselves as being present to ourselves, at times at least. To accommodate an entity with self-awareness or reflective experience within one's ontological domain is a challenge for any ontological enterprise. Proponents of a naturalist ontology, however, have a particularly difficult time to fit themselves into the framework they recommend – where to place their own experience within the worldview they endorse?

Let us assume that we are naturalists – so that we can savor the predicament we are in, and use the felt paradoxical tension to launch us into arguments. If we are naturalists, we will begin with two moves. First we will distinguish between metaphysics and ontology. We will say that naturalism is a claim about the metaphysical significance of scientific theories. We will emphasize that science is a distinctive form of interaction with 'reality' – in controlled experiment reality 'talks back' – it is most rational to trust the 'descriptiveness-in-principle' of scientific theories, even though any current scientific theory of 'reality' may be an inaccurate description. We might further expand on the phrase 'most rational' in the vein of Wilfrid Sellars' transcendental scientific realism; we might point out that we conceive of the praxis of science as a self-correcting enterprise and claim that it is only possible for us to entertain this conception if we assume that the relevant praxis is guided by *interactions* with reality. But no matter how we flesh out the metaphysical claim that it is most

rational to assume the descriptive potential of scientific theories and to take one's bearings from science, we will also emphasize that naturalism does not entail any specific ontology. The assumption that "science is the measure of all things, of what there is, that it is, and of what there is not, that it is not" – Sellars' tongue-in-cheek *scientia mensura* principle – does not prescribe which ontological categories we should use. The principle only introduces the constraint that the basic entity types postulated in the ontological description of a domain D be less informative than the basic notions of current scientific descriptions of D.

With this first move, then, we will try to downplay the difficulty: naturalism does not entail materialist ontologies, and even if a thorough-going reduction to physical entities were possible, the last word on the nature of physical entities is not spoken.

Our second move, if we are naturalists, will consist in a 'divide and conquer' strategy. We could again follow in Sellars' footsteps and say that there are two questions that we need to distinguish. The question of how to place sensory experience and its qualitative aspects into a naturalistic ontology should be separated from the question of how to account for intentionality or the aboutness of conscious thought. Once these two questions are separated, we will argue, we will see rather clearly that we can make sense of aboutness within a naturalist approach. In particular, we will see that in order to account for aboutness or cognitive content, we do not need the notion of representation – at least not in the strong semantic sense of the term that has led many philosophers to believe that mentality has features that do not fit into a naturalist metaphysics. With Sellars we will maintain that a cognitive content – e.g., a thought – is not a representation of something else, e.g., a mind-external state of affairs; rather, a cognitive content is the way in which a neurophysiological episode is functioning within its processual environment and such functioning can be described as a distinctive *type of processing* within a process system of a certain kind. Similarly, the so-called 'qualitative' aspect of sensory experiences – to the extent to which they are not *qualia* in the

sense of reflected qualities and thus cognitive contents – can be modelled as other distinctive types of processing within process systems that comprise goings-on within an organism and the organism's environment.

That we so far have not sufficiently appreciated this approach to a naturalist ontology of mind, we will explain, has to do with our fixation on static categories – as long as we try to cash out 'functionalism' in terms of *functions* rather than *functionings*, we miss out on all of its beauty. Similarly, as long as we do not take seriously the claim that the category of processes or 'ways of goings-on' contains what we need to model sensory 'qualia,' we are bound to mistake modeling for elimination and to join Daniel Dennett in "wondering where the yellow went." Against this we will stress that the category of process used in these descriptions of cognitive contents and sensory experience is *not* a 'material' item in the common sense but *a way of going-on*, or a dynamics of a certain kind. We will insist that this process-ontological account of sensory experience and cognitive content is not 'reductive' in the common sense of this term, since it does not define sensory experience and cognitive contents in terms of conditions of material objects.

And yet, even if we could work out in detail Sellars' process-ontological vision of how experience and thoughts fit into a naturalistic ontology – drawing on theories of complexity, self-organization, and embodied cognition – even if we had placed our sensory experiences and the contents of our thoughts, would we have succeeded in placing *ourselves* into the naturalist picture?

You are not present to yourself as some sensory and cognitive system – you are present to yourself as yourself. The real difficulty that we are up against resonates with the poetic grievances Whitehead brought against traditional metaphysics in 1929:

> All modern philosophy hinges round the difficulty of describing the world in terms of subject and predicate, substance and quality, particular and universal. The result always does violence to that immediate experience which we express in our actions, our hopes, our sympathies, our purposes… We find ourselves in a buzzing world, amid a

> democracy of fellow creatures; whereas, under some disguise or other, orthodox philosophy can only introduce us to solitary substances…

We may have moved, with process-ontological naturalism, beyond the traditional conception of humans as "substances," we may have put human organisms into a "buzzing world" – but one of quantum fields, not of fellow creatures. There is no hope, sympathy, or purpose in our naturalistic world of processings as long as we have no place for ourselves. In fact, no processing system in the envisaged naturalistic ontology could even be called solitary.

How, then, can we make room, within a naturalist ontology, for 'what it is like for a human being to be himself or herself'? We might begin by applying the naturalist interpretation of cognitive contents to the content of our self-experience. We might suggest that the cognitive content of 'I' is the functioning of the personal pronoun within a system of linguistic practices. Furthermore, we will try to treat the first person perspective of conscious experience in analogy to the naturalist analysis of sensory experience. In our process-ontological interpretation we transformed 'I am seeing something blue' into 'I am seeing-blue-ly'; now we will suggest to repeat the 'adverbialization move' and turn the subject of a visual experience into a first-person aspect that inherently belongs to the kind of going-on in the processing we call a seeing: 'seeing$_{\text{first-person}}$blue-ly'.

Can these strategies for naturalizing self-experience ultimately succeed? We can set the question aside, since even if they were, we still would not have brought into the picture what introspectively seems to matter most in self-experience: that we exist throughout time, our transtemporal unity. Without transtemporal unity, we are still a long way from hopes, sympathies, actions, and purposes.

However, a particular difficulty here is that our transtemporal unity is a form of dynamic continuity. Upon closer inspection, what we commonly call personal 'identity' upon closer inspection is a tapestry of unconscious and conscious acts by which we continuously build our self-conception. Staying with our process-ontological pre-

dilections we may suggest then that what we are to ourselves is the process by which we build and rebuild our self-conception.

To avoid succumbing to metaphors in this slippery domain of reflective subjectivity, we could spell out the notion of a self-conception more formally as a structural model of the complex process of conceiving of ourselves (a process that is cognitive in the broad sense of both sensory-experiential and conceptual). We could define a 'self-conception' as a selection function that ranges over concepts for capacities, where we would use 'capacities' as an umbrella term for – arbitrarily specific or generic – abilities, attributes, experiences, and statuses, e.g., *being a father, being a decision theorist, being creative, being athletic, having walked often through Schenley Park, being a member of the Royal Swedish Academy of Letters, working in Lund, being a Caucasian male* etc. A self-conception, we might suggest, is a selection function that maps such concepts of capacities onto identificatory narratives.

We could try to make this Leibnizian approach a bit more realistic and incorporate lessons of social determination. In defining the range of the selection functions we could postulate that any individual human being has (i) *sortal capacities* due to being a member of a certain kind, and (ii) *individual capacities*. Your 'personal identity' so-called is a selection function defined on your individual capacities (and experiences). Quite analogously, your 'cultural, social, and ethnic identities' so-called are selection functions defined on your sortal capacities. While you yourself determine the domain and outcome of the selection function for your personal capacities, the domain of the selection function for your sortal capacities is partly determined by the relevant cultural, social, and ethnic groups of individuals, not by yourself.

In more detail, let: $A = \{a_1, \ldots a_n\}$ be a group of individual human organisms, $T_{a1}$ be the set of (possibly overlapping) temporal intervals $t_i$ that together span the past and present lifetime of an individual $a_1$ in A, let $C_{a1\ t1}$ be the set of capacities of individual $a_1 \in A$ during $t_1 \in T_{a1}$. Let $C_{a1}$ be the union or total set of all such temporary capacities, $C_{a1} = \cup C_{a1,\ t1}$, and let us call these the 'diachronic capaci-

ties' of individual $a_1$ during any interval $t_i$. Let $V_{a1}$ be the powerset of $C_{a1}$. Finally, let $C_{ai}$ be the union of the diachronic capacities of all members of the group A, and $V_{ai}$ the set of possible group narratives for A.

The selection functions for (1) an individual's personal identificatory narratives and (2) various types of sortal identificatory narratives, i.e., cultural, social, or ethnic 'identity', can then be stated as follows:

> *An individual's personal self-conception* is a function that maps, for any given temporal interval $t_i$, a set of diachronic capacities (call this the individual's 'current introspective narrative') onto a set of diachronic capacities (call this the individual's 'projective narrative'). The introspective narrative and the projective narrative may be identical or overlapping (normally they will overlap a lot and only in pathological cases they will be discrete). The total set of argument-value pairs of this function are the individual's personal identificatory narrative.

Intuitively speaking, people select, for intervals of various durations, a collection of capacities in terms of which they wishes to define themselves during these intervals, and these projective identities depend on how they see themselves (during a stretch of time). (For those who prefer formulas, an individual's personal self-conception is the function $f_{pi}$: $T_{a1} \times C_{a1} \rightarrow V_{a1}$, with $f_{pi}(t_i, C_{a1}) = V_{a1,ti} \in V$). Even more briefly, an individual's personal narrative is a list of capacities that the individual considers to be representative for who she or he is during a spell of time, and endorses as valuative commitments.

> *An individual's sortal self-conception* is a function that maps, for any given interval of the group's lifetime, a set of sortal capacities (call this the individual's current sortal narrative) to a set of sortal capacities (call this the individual's projective sortal narrative). The total set of argument-value pairs of this function are the individual's sortal identificatory narrative.

Intuitively speaking, even though we are not entirely free to choose our sortal capacities, i.e., the identificatory capacities of the groups we belong to, we select, for some

time, which of the sortal capacities that characterize the group's members we wish to adopt as part of our projective narrative, and these projective selections depend again on how we see our sortal capacities currently. (To offer again a formal version as well, an individual's sortal self-conception is the function $f_{si}$: $T_{ai} \times C_{ai} \rightarrow V_{ai}$, with $f_{pi}(t_i, C_{ai}) = V_{ai,ti} \in V_{ai}$). Even more briefly, an individual's sortal narrative (i.e., cultural, social, ethnic narrative) is a list of capacities that the individual selects from a set of capacities that a relevant (cultural, social, ethnic) group considers to be representative for who they are during a certain interval; the individual considers the selected capacities as representative for herself or himself for a certain temporal interval and endorses them as valuative commitments.

An individual's full self-conception during a given interval, we could say, consists of the union of the relevant parts of his or her personal and sortal identificatory narratives. To receive a notion of the transtemporal unity that we are to ourselves, all we would need to do is to introduce a measure of variation for selection functions. A selection function's degree of internal variation would be defined via a similarity measure on the components of identificatory narratives. People who live in static communities have pleasantly (or depressingly?) monotonous sortal narratives – their sortal selves are stable. In contrast, people who live in fast-paced communities, or change communities due to emigration or career change, need to adjust frequently to altered sortal 'identities.' Similarly, our definition makes room for the full spectrum of variation in personal 'self-design', for the Indiana Jones' and the Immanual Kants among us – with reference to whether personal narratives are changed often or rarely, dramatically or minimally, we can distinguish between the volatile and the conservative, between those who frequently reinvent themselves completely, those who undergo partial conversions, those who allow for slow drifts in their self-conception, and those who doggedly reaffirm themselves.

But would all this descriptive machinery solve our predicament? By defining the notion of a self-conception in terms of selection functions and their degree of internal variation over time, we have arrived at a dynamic version

of what was classically called our 'empirical Ego.' And isn't this enough? Surely a description of our 'empirical Ego' is about all we are to ourselves?

If we are honest with ourselves and with our fellow philosophers, it is quite obvious I think that smartening up the empirical Ego with a bit of time-sensitive internal variation does not really get us to the heart of the phenomenology of self-experience. We all know, indubitably know, that we are in strange ways both *more and less* to ourselves than our 'empirical ego's' or self-conceptions. We all know, indubitably know, that what we are to ourselves lies way below the threshold of conceptualization and is at once poorer and richer in information.

To us naturalists this phenomenological realization comes as a shocking truth, especially when we face it in its most eloquent and perceptive presentation, in Henri Bergson's characterization of our immediate pre-conceptual self-awareness as "duration" (durée) or flow of consciousness. We experience the becoming that we are to ourselves as complex, says Bergson – yet there is no determinate structure of this complexity; we experience it as a multiplicity – yet not as a plurality where single elements could be distinguished; we experience it as a unity – yet also as continuous internal difference and change; we feel it as a paradoxical movement of turning towards as it turns away.

This, then, is the ultimate challenge for naturalism and our predicament in its final analysis: how can we accommodate the way in which we are present to ourselves, if this 'way' or mode of being apparently defies all conceptual articulation? We cannot dodge the phenomenological datum that we are aware of ourselves as Bergson's 'durée, as a dynamic multiplicity of inchoate, potential contentualities that are not yet brought into the discreteness of conceptual articulation. What you are to yourself is like a complex melody that you feel 'synthetically' but cannot, as such, hear 'analytically' as a sequence of tones and chords. Or if you do, you lose the original phenomenon.

Speaking with the history of philosophy, we use our 'empirical Ego' or identificatory narratives for practical purposes of self-declaration, but we are to ourselves the 'transcendental Ego' – which, as Bergson saw first, is by no

means the empty self-identical 'I' but an internally variegated, rich, dynamic affair *below* the threshold of theoretical articulation.

What can we do in order to bring the becoming that we are to ourselves *somehow* into the domain of theory in general, and into the scope of a naturalist ontology in particular? We need to capitalize on the fact that processes come in different dynamic shapes, and that these difference matter for how we can refer to temporal parts of the process. Activities, for example, such as *singing* or *reading* or *snowing* are monotonous or dynamically homogeneous processes. They do not have distinctive phases, i.e., pretty much every temporal part of an activity is like any other and like the whole activity (I pass over the finer points to be made concerning the likepartedness of activities). Developments, on the other hand, such as a *baptism* or *a caterpillar's turning into a butterfly* or *DNA-replication* or *vinegar reacting with baking soda* have temporal parts that differ in kind, forming separable phases. Once we have established a kind term for a development (e.g., *a human life*, *an explosion*, *making lasagne*), we can refer to a phase of a development of this kind in various ways. Sometimes the phases of a development have kind tems of their own (e.g., *childhood*, *autumn*, *matchball*). Often, however, we perform a synecdoche, and refer to the phase in terms of the whole development, *toto pro parte* (e.g., *the finish of the marathon*) or in terms of another phase of the development, *pars pro parte* (e.g. *the first indications of his later success*).

Applying this observation to the task at hand, in order to put our pre-conceptual self-experience into naturalist ontology, we might try to capitalize on the 'observer effect' in self-comprehension – the fact that any attempt to conceive of what we feel we are to ourselves destroys the phenomenon. J.J. Gibson's category of "affordance" can be usefully applied here, if we read it as referring to the beginning of a development that we conceptualize via its end. The flow that we are to ourselves is that which affords a self-conception, we might say – at times and across times. With reference to our definition of self-conception above in terms of selection functions, we could postulate that the flow that we are to ourselves is that

which affords each of the values of both selection functions, personal and sortal. Changing our terminology slightly, we could say that each of the values of a selection function is a set of capacity ascriptions that are the 'endpoints' of developments afforded by the flow of our self-awareness at that time.

As such this provides us with nothing more than an indirect reference to the beginnings of processes of conceptualization, where the latter can be further specified as developmental processes (as opposed to activities) and of the kind of conceptualizations that occur within us when we try to comprehend ourselves. We would need to set out in greater detail the kind of link that is introduced by calling something an 'affordance.' Some authors have tried to define affordances as dispositions, but this misses Gibson's points entirely, in my view. To say that hazelnut A affords nourishment B for squirrel C by way of a digestion D (of A in C) is not to refer to a disposition of the nut, which could be as short-lived as A's involvement in D; rather, it is a reference to a correlativity between A and B that is (i) established by the evolutionary history of items of type (species) C and (ii) realizable in D. So affordances are best understood, in my view, as the first stages of developments that are rendered frequent, and thus identifiable as a kind, due to an embedding feedback dynamics.

Assume then that we had a satisfactory notion of affordance and a plausible definition of a self-conception across a temporal interval, would we have succeeded in putting ourselves into the naturalistic picture if our pre-conceptual self-experience is that which affords our self-conception across time, allowing perhaps also for a *modicum* of emergence in the definition of affordance? Almost.

We would (at best) have gotten hold of the contentual indefiniteness of the pre-conceptual experience of ourselves, but we still would not have captured the flow, the becoming that we are to ourselves. What could we do to bring 'dynamicity' into the picture?

Our previous move provides the basic strategy. The history of ontology – although largely in the grip of what I call the "myth of substance," a fixation onto the description of static Being and beings – also features various

attempts to conceptualize dynamicity. A preferred option appears to be to approach dynamicity via self-reference, self-propagation or self-realization: as the producing of P, where P is precisely this very same producing; self-propagation requires that we read the expression 'this very same' as 'the same kind,' self-realization requires that we read the expression 'this very same' as a 'the same individual becoming,' whose individuality is defined as continuously incomplete or in the making. Aristotle's notion of *physis* still provides the most detailed ontological description of either self-propagation or self-realization (scholars are divided on whether your physis is a kind of activity or an individual activity), and recent semantics for self-referential sentences (e.g., circular definition), as, e.g., the one used in the "revision theory of truth" by A. Gupta and N. Belnap, can reassure ontologists that there is nothing *per se* disreputable about self-referential expressions.

We might say, then, that the flow that we are to ourselves is that which affords our conceptualizations of dynamicity in terms of self-realization. The becoming of our selves is a branching development that produces (affords) self-conceptions on the one hand, and the idea of dynamicity on the other hand.

Have we naturalists thus worked ourselves out of our quandary? Or are the preceding ruminations more a *reductio* of naturalism? Keeping in mind Whitehead's remark that all philosophy is "footnotes to Plato," I have used the subjunctive wherever possible to begin a footnote on *Sophist* 260a: "logos [reality] is born for us through the weaving together of forms," Plato says ambiguously, leaving us without comment on whether the weaving is ever within conceptual reach or must simply be lived.

# Confronting the collapse of humanitarian values in foreignpolicy decision making

PAUL SLOVIC, ROBIN GREGORY, DAVID FRANK, AND DANIEL VASTFJALL

## 1. *Introduction*

Nils-Eric Sahlin's broad understanding of values, preferences, and decision making has led him to engage psychologists, economists, philosophers, and others in stimulating discussions of rationality. In a 2010 paper, "Ethical theory and the philosophy of risk: first thoughts" (Journal of Risk Research), Johan Brannmark and Nils-Eric argue that evidence from a half-century of psychological research paints a picture of human beings

> possessing a virtual tool-box of heuristic devices, affects and instincts. These devices might be perfectly explainable by evolutionary processes; they perform well in some contexts and lead us into error in others. Given such a picture, irrationality involving violations of classic axioms of rationality will be an endemic and ineliminable feature of human behaviour. This does not necessarily mean that there is anything wrong with us. What it does mean is that the background conception of human agency we should have in mind when thinking about normative ethics is not so much Rational Man as something like Heuristic Man: a person whose competence is context-dependent and whose deliberative skills do not necessarily transfer from one context to another; a person whose modes of thinking will inevitably be deeply shaped by a range of factors beyond agency as such.

From this conclusion, Brannmark and Sahlin argue for a new kind of theory to be applied to moral issues based on the notion that "if human decision-making is ultimately fragmented, in the sense that we have a diversity of concrete and often disunited ways of coping with different types of situation, then perhaps normative ethics should not be working towards unified and all-encompassing theories". They propose a "mid-level" theory of rationality to remedy their view that "large areas of common-sense morality and current ethical practice implicitly rely on outmoded pictures of human agency, and that reform would in fact greatly improve them".

The present essay sketches a portrait of top-level decision making in the face of moral crises that we believe fits well within the framework proposed by Brannmark and Sahlin. It draws on decades of old research on values and preferences to explain what appears to be a systematic bias leading to the discounting of humanitarian objectives in major foreign-policy decisions.

## 2. *The problem and a hypothesis*

Decisions to save civilian lives by intervening in foreign countries are some of the most difficult and controversial choices facing national decision makers. Although each situation is unique, decisions involving tradeoffs that pit the value of human lives against other important objectives are quite common. For example, in 2011 the United States supported military action to protect the lives of civilians living in Libya and recently intervened aggressively to protect a threatened population of Yazidi people in Iraq. On the other hand, the United States has done little to intervene in the genocide in Darfur or the mass atrocities in Syria that have led to hundreds of thousands of deaths and millions of displaced persons.

One explanation is that the threat to lives in Darfur or Syria has not been valued highly enough to compete against other political, economic, cultural, or military objectives. However, because these decisions typically are made at the highest levels of government and without transparency, very little is known about the discussions

and debates that take place or the relative importance placed on different goals and concerns. What we do know, however, is that there is often a striking disconnect between the high value placed on saving human lives expressed by top government officials and the apparent low value revealed by government inaction when millions are threatened.

On the basis of theoretical models of judgment and choice, research in social cognition, and careful reading of official pronouncements, we have developed a hypothesis to explain this disconnect. We believe that multiple objectives are in play and that highly regarded humanitarian values essentially collapse in the competition with national security and economic security objectives.

Thus, in situations where the United States has intervened with the stated objective of saving lives, there were presumed security benefits as well. Libyan leader, Moammar Gadhafi had long been known as a loose cannon, addicted to violence at home and elsewhere. His menacing visage adorned the cover of Time magazine four times since 1986, when Ronald Reagan referred to him as "this mad dog of the Middle East." Similarly, security objectives were important in Iraq. In addition to protecting the Yazidis, we were protecting American military and diplomatic personnel stationed nearby in Erbil.

In contrast, humanitarian intervention in Darfur and Syria has posed threats to security. Omar al-Bashir, who takes a back seat to no one as a murderer, had been providing information about terrorist activities to the American government. Moreover, the Chinese have been a strong protector of the Sudanese regime and action against al-Bashir that strained relations with China would threaten the United States' economic and military interests. Similarly, American humanitarian intervention in Syria might be seen as an aggressive political act posing many threats to U.S. security. But the rise of ISIS, itself a security threat and a humanitarian threat, has led to intervention in Syria. In sum, these examples suggest a hypothesis to the effect that the U.S. only seems to launch humanitarian interventions when security interests are also served by such action.

Tradeoffs between security and humanitarian objectives are nearly always made implicitly, as part of the decision making process. As a result, it may be that this disconnect between our lofty stated humanitarian values and their disregard as revealed by inaction is not consciously recognized by the decision makers as they debate—typically behind closed doors—competing and often complex decision options.

### *3. Theoretical underpinning for the hypothesis*

Economists, philosophers, and other students of choice have long been interested in the influence on decisions of expressed or stated values as compared to values that are revealed through choices. Rational choice theories typically assume that choices are consistent with expressed values. However, a great deal of empirical research has shown that the values indicated by these two modes of assessment often differ. One explanation for such inconsistency has centered around the weighting of the various attributes or objectives of decision options and the evidence for systematic discrepancies in weighting associated with expressed and revealed preferences.

For example, a study by Slovic published in 1975 found that difficult choices were systematically decided in favor of the alternative that was superior on the most important attribute. In 1988, Tversky, Sattath, and Slovic used this finding as a springboard to a general theory of choice called the "contingent weighting model." At the heart of this model was the "prominence effect," which recognized that the more prominent attributes were weighted more heavily in choice than in judgments reflecting stated preferences or values. The presumed explanation for this effect is that, unlike stated values, chosen actions need to be justified.

We argue that the prominence effect may underlie the observed disconnect between expressed and revealed values regarding whether or not to act to save large numbers of civilian lives under attack in foreign countries. Specifically, we hypothesize that national security is the prominent dimension in the context we are studying here. Chosen actions need to be justified, and deciding in favor of secu-

rity likely makes a stronger argument than deciding in favor of saving foreign lives, no matter how many thousands or millions of lives are at stake.

In the political and military arenas, there are many examples of security objectives appearing to override important rights such as those associated with civil liberties. Academic support for the prominence of security in personal behaviors comes from an Israeli study by Shnabel and colleagues, who found that people first seek to satisfy their needs for safety and security and only then do they authorize themselves to seek the satisfaction of higher-order needs including maintaining a positive moral image and social relatedness with others. Similarly, Mikulincer et al found that people who have secure social attachments find it easier to perceive and respond to other people's suffering.

We are just beginning to conduct laboratory experiments to determine the impact of security prominence in scenarios based on the humanitarian crisis in Syria, where the objective of protecting 100,000 civilian lives is pitted against the political and military risks of American intervention. Preliminary results support the hypothesis that an individual's strong expressed values for intervening to protect lives are often contravened by that same person's decisions in favor of non-intervention.

## *4. Concluding remarks*

The implications of the prominence effect for theories of choice are challenging. As Tversky et al observe: "If different elicitation procedures produce different orderings of options, how can preferences and values be defined? And in what sense do they exist? [...] In the absence of well-defined preferences, the foundations of choice theory and decision analysis are called into question".

Beyond the theoretical significance, if it is true that prominence devalues efforts to intervene in massive humanitarian crises, the moral, ethical, and strategic implications of this bias may be profound. Whether structured value elicitation procedures, informed by an understanding of the prominence effect, can bring stated and revealed values into alignment, remains to be determined.

# Det eviga livet

PETER SYLWAN

Man skall vara rädd om sina gener – någon annan kan behöva dem. Förutom barnen alltså. Jo – jag vet att jag passerar gränsen. Så här får man inte säga och spekulera i en festskrift med anspråk på att vara vetenskaplig. Men jag kan inte låta bli. Historien är för bra för att vara sann. Men jag tänker inte kolla. Det skall man aldrig göra med en bra historia – då kanske den spricker. Och då blir tankar som inte skulle ha tänkts aldrig tänkta, experiment som inte skulle ha gjorts aldrig gjorda och upptäckter som inte skulle ha upptäckts aldrig upptäckta. Dessutom är den här historien lätt att snyta ur näsan – ni skall strax förstå varför och på vilket sätt. Men så här ligger det till.

Det kom ett pressmeddelande från Köpenhamns Universitet om en upptäckt publicerad i PNAS (de måste väl ändå ha kollat) Ett danskt/norskt team av genjägare hade hittat en 43 000 år gammal mammut – som levde. Inte en hel förstås. Bara en så liten del att den rymdes i en bakterie. De hade hittat små fragment av 43 000 år gammalt mammut-DNA – det var till och med av den ulliga typen – inbyggt och fungerande inne i arvsmassan på en nu levande bakterie! Man häpnar. Tala om återfödelse eller åtminstone återbruk. Genjägarna själva tror att bakterier kan använda DNA-stumpar som är flera hundra tusen år gamla. De bara finns där ute – mikroberna. Bakterier som simmar, kryper och flyger omkring bland en massa uttjänt och okänt DNA som gamla skrothandlare och kittelflickare som plockar på sig av vad som finns och använder det som duger. Tänk om bakterier inte ens finns? Alltså inte finns var och en som sin sort med bestämda egenskaper. Att de istället uppstår – var och en efter sin art när det

är läge och plockar på sig det DNA de behöver för tillfället för att skaffa sig de egenskaper de vill ha och får dem att överleva. Och det finns en del att ta av. Det är nu jag kommer till näsan.

Var försiktig när ni nyser eller snyter er. Man kan aldrig veta vad som snuvan består av. Inte än i vart fall. Förkylda genjägare som dyker på djupet i snuvan kommer upp till ytan med en hel ännu okänd DNA-värld. Mer än 20 % av allt DNA de hittar i näsan kommer från fullständigt okända organismer. Det är en obehaglig tanke. Att min näsa är befolkad av livsformer – eller åtminstone DNA som ingen haft någon aning om hur de ser ut eller varifrån de kommer. Det är likadant i magen – eller på huden, i blodet, i jorden och haven. Var än genjägarna jagar så är det samma sak. Bytet de tar med sig hem till labbet visar sig innehålla DNA som till ibland ända upp till 90 % kommer från helt okända organismer – eller helt okända livsformer. Vad säger en forskare som möter en helt okänd livsform? Det går ju inte med ”dr Livingstone I presume”. Om honom visste ju Stanley åtminstone att han gett sig ut i djungeln. Men om de här livsformerna vet vi ju ingenting ännu – mer än att de kanske finns eller uppstår när det är läge.

En av de märkligaste mikroorganismer jag överhuvudtaget har hört talas om talar för att det går. Den kommer till och med alldeles själv tillbaka från döden. För bakterien *deinococcus radiodurans* duger inga som helst vanliga dödsbegrepp. Den kan vara dödare än Monty Pythons döda papegoja och ändå leva upp igen. Sätt ut *radiodurans* i öknen. Låt solens UV-strålar bränna den till döds. Begrav den i kärnkraftsavfall. Bestråla den med radioaktivet 100 gånger starkare än allt liv står ut med. Spräng dess DNA i tusentals småbitar. Det bekommer den inte. Den är visserligen stendöd och visar inte minsta tecken på liv. Men när solen går ner, radioaktiviteten försvinner, vattnet kommer tillbaka – då sopar den ihop sitt DNA igen och sätter samman bitarna till fungerande gener och börjar leva igen.

Tänk om det är så det går till över huvud taget? Att det enda som finns för evigt är en gränslös DNA-soppa av sopor som lagar sig själv efter läge? Styrda av slump och nödvändighet. Tillfälligheternas spel och tidens sorterings-

verk. Och när det blev läge för oss så fanns vi där plötsligt. Skaparen är skrothandlare och människan en sopa? Plötsligt och plötsligt förresten. Det har ju tagit några miljarder år för oss att uppstå ur den globala soppan av ur-DNA. Men vad är det mot evigheten. Det gäller att ha perspektiv och tålamod. Och gillar man inte sopor så är ljusgestalter kanske trevligare. Det var ju så det började. Att allt blev ljus efter Den Stora Smällen. Resten är ju historia och ljus förvandlad till materia enligt Einstein och $E=mc^2$.

Världen som stjärnstoft och människan som inkarnerat – och besjälat – ljus. Det är bättre än sopor. Jag ser gärna livet som en evigt pulserande och varierande ström av DNA som hela tiden formar sig till olika virvlar av olika livsformer som kommer och går. Också vi är bara virvlar i DNA-havet – ”ett bloss i vind. Ett födsloskri och en fårad kind” – för att tala ferlinska. En ljuslåga av någorlunda bestående form och funktion i några korta decennier trots att vi egentligen består av en ström långsamt brinnande kol. Den yttre formen och den inre lågan är densamma – eller förändras bara långsamt. Trots att nästan alla våra celler byts ut flera gånger under en livstid är vi ändå samma personer sammanhållen av och fritt svängande i våra geners spiralvridna repstege. Och en del av det som gör oss till dem vi är lever kanske vidare så länge DNA består. Så man skall vara rädd om sitt DNA. Någon kanske behöver det i nästa liv. Virus till exempel.

Liv och liv förresten. Det är väl osäkert om man skall kalla virus för liv. De åker ju bara snålskjuts på allt och alla som lever på riktigt. Smyger in sin egen arvsmassa bland våra gener, plockar på sig vad som kan vara bra att ha, lämnar kvar en del som de kan vara utan och far vidare till nästa liv. Bortåt 8 % av hela vårt DNA kommer från virus. Det är ungefär i klass med vad vi använder till våra egna gener. Lika mycket vi som virus! Och det är vi själva som bär dem vidare från generation till generation i evigheter. Det är smart. Tala om fripassagerare i tiden. En del av det virus-DNA vi bär på och för över till våra barn är 100-tals miljoner år gammalt. Vem sa något om evigt liv? Den ulliga mammuten har inte mycket att sätt emot viruset i den tävlingen. Mycket talar för att virus är ett av evolutionens viktigaste verktyg för att sopa ihop och flytta runt de

DNA-stumpar som behövs för att åstadkomma variation och förnyelse. Utan virus inget vi. Inte underligt kanske att av allt DNA som överhuvudtaget finns i världen så är den allra största delen – just det – virus. Vill man dela med sig till gemenskapen och evigheten är det bäst att välja jordbegravning. Eller möjligen vacuumfrysning, pulverisering och bli borta med vinden. Men aldrig aska. Mitt DNA skall inte gå upp i rök – det känns lite oansvarigt. Någon annan kan behöva det. Och vem vet – kanske kommer en liten bit tillbaka till livet igen i en svala eller annan människa om 40 000 år.

Fast vill det sig illa kanske jag hamnar i en livsform som jag nog inte gärna vill ha med att göra. Naturen gör som Kajsa Varg och tager vad den haver. Genteknikens kritiker som ser själva genflyttandet som något onaturligt och just därför olämpligt i sig har missat en naturlig poäng. Den är i sig tämligen promiskuös och inte alls så noga med gränserna. Genteknikernas alla verktyg har kunnat hämtas från naturen just därför att naturen själv är den mest avancerade och sofistikerade manipulatören. Och det gäller inte bara gener. Det gäller också hjärnor. Att hamna som DNA-leverantör till toxoplasmi gondii (TG) är inget jag vill ha i mitt CV för evigheten – möjligen som ett verktyg i verkligheten. TG lever dubbelliv. Ett liv i råttor och ett i katter. Det ena kräver det andra. Och när TG lever i råttan – eller musen – invaderar den hyresvärdens hjärna och får hen att se katten i ett helt nytt ljus. Det verkar nästan som om en TG-råtta eller mus söker sig till katten. Och när katten ätit upp sin middagsmat åker TG:s ägg ut med kattskiten och hamnar så småningom i en musmage igen. En annan variant på samma tema lever dubbelliv bland myror och mular. Det normala för en myra mot kvällningen är att gå hem och lägga sig i stacken med de andra myrorna. Men de som haft oturen att få i sig en liten larv av lilla leverflundran kommer på helt andra tankar. Mot kvällningen gör den myran inget annat än längtar efter utsikten från ett grässtrå. Väl där biter den sig fast i gräsets topp. Under natten och morgontimmarnas månsken och morgondagg återgår lilla leverflundran med myrans och mularnas hjälp i sitt eviga kretslopp. Man kan bara hoppas att myran och musen var lyckliga.

Tänk vad lite vi vet. Eller visste. Eller vad lite vi vet om det vi inte vet. Vetenskapsjournalistikens vanligaste fras är ”än vad vi tidigare trott”. Den ger över 12 miljoner träffar på nätet. Preciserar man den till att gälla bara vad forskare tidigare trott blir träffarna över 1 miljon. Det är därför man aldrig kan lita på forskare. De ändrar sig ju alltid. Men det är ju själva vitsen med vetenskapen. Att hitta nya kunskaper som visar att de gamla var fel – eller åtminstone ofullständiga. Det är därför man kan lita på vetenskapen. Området vi fattar att vi inte fattar något om blir ju bara större ju mer forskarna forskar. Om det vi inte vet kan vi säga samma sak som ärkebiskopen och muslimer säger om Gud. Att det är större – alltid. Hur mycket man än försöker förstå. Förr visste ju alla att jorden var platt och tog slut vid horisonten. Nu när vi umgås nästan dagligen med Higgs har vi förstått att allt som vi vet något om till 95% – eller mer – har vi inte en aning om – mörk energi och mörk materia. Med det mörka universum är det som med generna, jorden och DNA:t man snyter ur näsan. Det mesta finns kvar att upptäcka. Och ju mer vi förstår desto större blir området vi fattar att vi inte vet något om.

# Chance, love and logic: Ramsey and Peirce on norms, rationality and the conduct of life

CLAUDINE TIERCELIN

> . . . The highest ideal would be always to have a true opinion and be certain of it; but this ideal is more suited to God than to man (F.P. Ramsey, Truth and Probability, 89–90)
>
> We must not begin by talking of pure ideas, – vagabond thoughts that tramp the public roads without any human inhabitation, – but must begin with men and their conversation. (Peirce, CP 8.112.)

Ramsey's pragmatism has now received some attention (Engel 1983, Sahlin 1990, Levi 1997, Hookway 2000, Dokic & Engel, 2002, Tiercelin 1993a, 2004d, 2005). But the "essence" of it is still fuzzy: Ramsey claimed that "his pragmatism was derived from Mr. Russell and was of course very vague and undeveloped" (ibid.). He recognized his "indebtness to Mr Wittgenstein, from which [his] view of logic was derived" (PP, 51), but also excepted from what was due to him "the parts which have a pragmatist tendency", which seemed precisely "to be needed in order to fill in a gap in [Wittgenstein's] system"(ibid.). Quite convincingly, it has been argued also that Wittgenstein's later work has a "pragmatist" flavour, maybe due, in part, to some impact of Ramsey on his work around 1930 (Sahlin 1990, 227, Hookway 2000, 136). Ramsey himself takes his "pragmatism" to be "that the meaning of a sentence is to be defined by reference to the actions to which

asserting it would lead, or more vaguely still, by its possible causes and effects" (PP, 51), before adding: "Of this I feel certain, but of nothing more definite".

Despite such uncertainties, it is rather clear too that many Ramseyan themes are much inspired by the founder of pragmatism, C. S. Peirce, who is referred to in several places, and whom Ramsey had been mostly acquainted with through the Cohen edition of *Chance, Love and Logic* (Peirce 1923), but also, as Galavotti (1991,16) observed, through a transcription-summary made by C. K. Ogden (Document 007-05-01 of the Ramsey Collection) of an important although never quoted paper by Peirce in his published writings, namely "Prolegomena to an apology for pragmaticism" (*The Monist* 16 (1906), 492–546, repr. In Peirce 1931–58, vol. 4, par. 530 to 572 (references to this edition hereafter quoted CP, then by volume and paragraph numbers).

As concerns truth, action, knowledge and inquiry (Misak 1991), induction (Levi 1980a, 1980b, 1997), probability, Ramsey's pragmatism has been shown to be close to Peirce's own version. Less obvious and yet, in my view, no less important aspects pertain to Ramsey's "mild" realism as regards the problem of universals, at least, up to a point, and provided realism is seen through the scholastic, semantic and yet scientific and metaphysical lens of Peirce's – rather idiosyncratic – own approach (Tiercelin 2004d).

In what follows, I would like to pursue along such lines and, focusing more this time, although briefly, on their respective views on logic and ethics, but also on norms, rationality, chance, and the conduct of life, I will try to show how close they were indeed.

## *1. Logic versus ethics. The anti-theoretical reaction*

Just as most pragmatists (James, Peirce, but also Wittgenstein), Ramsey was rather hostile to moral rationalism and had towards ethics, a basically anti-theoretical reaction (Tiercelin 1994, 2002b, 2004a, 2005, 2014):

> Any attempt, he writes, to treat such topics (of 'popular philosophy' such as 'the relation of man to nature, and the meaning of morality') seriously reduces them to questions either of science or of technical philosophy, or results more immediately in perceiving them to be nonsensical. [...] Theology and Absolute Ethics are two famous subjects which we have realised to have no real objects. (Epilogue 1925, 246–247)

In a similar vein, Peirce claims that "Ethics is not Practics": "it has no essential state". As "the science of the end and aim of life", it is excluded from philosophy, for its being exclusively "psychical", thus confined to a special department of experience, while philosophy "studies experience in its universal characteristics." (1992, 115–116). It ranks "with the arts, or rather with the theories of the arts", which "of all theoretical sciences" Peirce regards as "the most concrete", while philosophy is "the most abstract of all the real sciences". Hence no confusion should be made between Vital questions in which instinct, or the feeling of some "primitive obligation" should be followed (and where one should favour a form of Conservative Sentimentalism), and Scientific questions, in which, in strict parlance, belief – as a disposition to act – has no place (CP 1.655, 5.60):

> It is the instincts, the sentiments, that make the substance of the soul. Cognition is only its surface, its locus of contact with what is external to it. [...] Thus, pure theoretical knowledge, or science, has nothing directly to say concerning practical matters, and nothing even applicable at all to vital crises. Theory is applicable to minor practical affairs; but matters of vital importance must be left to sentiment, that is, to instinct. (1992, 110–112)

Indeed, one should "not allow to sentiment or instinct any weight whatsoever in theoretical matters, not the slightest" (1992, 111), for to make knowledge an adjunction to ethics, or to allow the intrusion of moral or vital factors in science, always makes one run the risk of judging the validity of reasonings according to the impressions they make on oneself: "When men begin to rationalize about their conduct, the first effect is to deliver them over to

their passions and produce the most frightful demoralization, especially in sexual matters."

> Men many times fancy that they act from reason when, in point of fact, the reasons they attribute to themselves are nothing but excuses which unconscious instinct invents to satisfy the teasing "whys" of the ego. The extent of this self-delusion is such as to render philosophical rationalism a farce." (1992, 111)

Men "continue to tell themselves they regulate their conduct by reason; but they learn to look forward and see what conclusions a given method will lead to before they give their adhesion to it." It is then the reign of "sham reasoning", when "it is no longer the reasoning which determines what the conclusion shall be, but it is the conclusion which determines what the reasoning shall be." (CP 1.57) "The effect of mixing speculative inquiry with questions of conduct results finally in a sort of make-believe reasoning which deceives itself in regard to its real character" (CP 1.56); even worse: "men come to look upon reasoning as mainly decorative." (CP 1.58). It is crucial to dissociate theoretical from vital matters, for ethics requires beliefs, and mainly firm and fixed ones. Now, the unavoidable dogmatism, conservatism, – "morality is essentially conservative"(CP 1.50) – urgency, required by this is incompatible with the disinterest, humility, sense of doubt and probability, uncertainty, refusal of Manichean distinctions, respect of fine nuances that are so characteristic of the theoretical (scientific) domain: experience, scientifically conducted, can never reach absolute certainty, exactitude, necessity or universality, whereas moral conscience requires uniformity, regularity, repeatability, as Robert Musil insisted on. This is why, as Peirce notes, "in more ways than one, an exaggerated regard for morality is unfavorable to scientific progress", and "as a means to good life, it is not necessarily coextensive with good conduct." (CP 1.50)

This is incompatible with the scientific attitude which Peirce defines less as a corpus of established truths and pieces of knowledge than as a mode of life, a pursuit of knowledge rather than knowledge. In order to be a real

scientist, philosopher, or more generally, inquirer, it is necessary to have "such virtues as intellectual honesty and sincerity and a real love of truth" (CP 2.82). The first step towards finding out being to acknowledge that you do not satisfactorily know already (CP 1.13). This is why a real scientist is not, strictly speaking, a believer, since belief is something upon which a man is willing to act (CP 1.635). He has only hypotheses, which are believed to the only extent that the economy of research prescribes, for the time being, that they should not be doubted, and that on them inquiry shall cease (CP 5.589); but they are revisable, hence provisional. For one should be ready to overthrow one's whole cartload of beliefs, as soon as experience requires it. Hence, there is nothing common between the man of science, moved by "a hearty and active desire to learn what is true" and "penetrated with a sense of the unsatisfactoriness of his present condition of knowledge" (CP 5.582) and the theologian or professor who is only moved by the desire to stick to his previous beliefs. Such antithetical attitudes are illustrated by Peirce under the two anti-scientific attitudes of fixing belief embodied by the methods of tenacity and authority, and under the two figures of the theologian (or Hegelian seminarist) and the scientist (CP 1.40). For Peirce, no compromise of principle is possible between science and society, morality or practice. Once for all, one should get rid of that "Hellenistic tendency" to "mingle Philosophy and Practice".

## 2. *Peirce and Ramsey on the need for normative sciences*

However, such a condemnation of moral rationalism does not prevent either Peirce or Ramsey from having a strict conception of rationality as a norm; and both emphasize the necessity to build a "doctrine of the normative sciences": on the contrary, it calls for it and for something that might transcend the mere laws of "association" and utility, so as to be guided by ends and ideals. For even if both should never be totally separated, it is an error, Peirce says, "to confound an ideal of conduct with a motive to action" (CP 1.574).

In the same way, Ramsey notes that Logic, Aesthetics, and Ethics "have a peculiar position among the sciences: whereas all other sciences are concerned with the description and explanation of what happens, these three normative studies aim not at description but at criticism." Indeed, although the trial of such critical disciplines "correspond to the three called fundamental values, truth, beauty, and goodness",

> … the correspondence, is by no means, exact. For whereas the chief question in Ethics is undoubtedly "what is good?", and in Aesthetics "what is beautiful?", the question "what is true?" is one which all the sciences answer, each in its own domain, and in no way the peculiar concern of Logic. (OT, 3)

All three sciences have something in common, namely

> … to account for our actual conduct is the duty of the psychologist; the logician, the critic, and the moralist tell us not how we do but how we ought to think, feel, and act. (OT, 3)

Such a conception is very close to the "doctrine of the normative sciences", worked out at length by C.S. Peirce (Tiercelin 1993b, 335–384), and part and parcel of science itself (CP 5.39). Although all three sciences are positive in so far as the assertions they make (in logic, ethics or aesthetics) rest on facts of experience which force themselves upon us (CP 5.120), they are not *practical* sciences, because their object is analysis and definition. So they are the purely theoretical sciences of purpose, of purely theoretical purpose (CP 1.282). They are the sciences of the laws of the conformity of things to ends. "Aesthetics considers those things whose ends lie in action, and logic those things whose end is to represent something" (CP 5.129). The new and important fact then is the definition of logic as a normative science, and even more, as a particular problem of ethics, which is in turn, dependent upon aesthetics (CP 2.197). Indeed, the essential problem of ethics is not right or wrong, but "what I am deliberately ready to accept, as the statement of what I want to do"(CP 2.198).

So it is mostly a science of ends. Thus, logic depends on it, since it has to do with thinking as a deliberate activity and with the means to reach that end, which is a valid well conducted reasoning. Hence, it becomes "impossible to be completely and rationally logical except on an ethical basis" (CP 2.198). But in turn, both depend on aesthetics, which is the analysis of the end itself, and of the ideal one would be willing to accept and to conform to (Tiercelin 1997, 41–42).

Just like Peirce, Ramsey is eager to stress the importance of the three normative sciences, and to view Logic as one of them; but he also emphasizes that "what Logic studies", more specifically, "is not so much the truth of opinions, as the reasonableness of arguments or inferences." And the distinction, he adds, "is an important one", which lies in this that "truth is an attribute of opinions, statements, or propositions", in first approximation, "accordance with fact." Hence, "if we have an opinion or statement by itself the most important point of view from which we can criticize it is that of truth and falsity, and the proper person to do this is not the logician but the expert on the particular matter with which the statement deals." However, one should be aware that

> … opinions and statements generally occur not by themselves, but as the result of some mental process, such as perception, memory, inference, or guessing. *It is precisely the logician's business to be concerned with the particular method of forming opinions known as inference or argument, and the inferences he approves of are not so appropriately called "true", and "valid", but "sound", or "rational"* … the primary subject of the logician is inferences or arguments, not opinions or statements, and his predicate of value is rationality not truth. (OT, 3–4, *my emphasis*)

Peirce gives a very close presentation of what Logic is mainly concerned with: if logic deals, for Ramsey, with "the reasonableness of arguments or inferences", it is, for Peirce, "the theory of the establishment of stable beliefs" and "the theory of deliberate thought", and reasoning itself (in order to be called, strictly speaking, an "inference" rather than a mere "argument") should be taken as

"thinking in a controlled and deliberate way" (CP 1.573), so true it is that for a pragmatist, the way one thinks cannot be distinguished from the way one conducts oneself (CP 5.534), thus from the way one is guided by a purpose or an ideal (CP 1.573), namely, that of the discovery of reality, which explains why, for Peirce as for Ramsey, "since the whole purpose of argument is to arrive at truth, there must be some relation between the soundness of arguments and the truth of opinions", even if, Ramsey admits "it is not easy to say exactly what the relation is." (OT, 3).

It is maybe in the last part of the sentence that we may find some difference between Peirce and Ramsey. While the former gives a very detailed analysis of the psychological mechanisms taking place in reasoning and inference, the latter, as P. Suppes noted, "despite his emphasis on the necessity of having a psychological method of measuring belief, in order to have a usable measurement"(2004, 36), "does not give many details about the genuine psychology going on in the process of reasoning itself" and does "not venture into this territory" (ibid., 52), maybe partly because "he was much too caught up in the writings of those at or close to Cambridge, of subjective probability." (ibid., 37), but also, somewhat surprisingly, since "Ramsey insists much, both in this subjectivist theory of probability and his theory of partial belief, and [in his important article on the foundations of mathematics (1931)] here, on the psychological dimension to be accounted for by his 'human logic of truth', supposed to be a much needed addition to the 'logic of consistency'"(ibid.): "Before coming to his real point; the logician is bound to begin by preliminary investigations into the nature and forms of opinions and statements, which must be conceded to belong properly to psychology since they are concerned not with values but with the actual characters of mental processes. Since, however, psychologists grossly neglect the aspects of their subject which are most important to the logician, they are commonly regarded as belonging to logic, and logic as the term is ordinarily used consists to a great extent of psychology. In the same way, students of ethics and aesthetics are obliged to undertake for themselves all sorts of psychological preliminaries." (OT, 4).

Now, not only did Peirce see, as Ramsey clearly did too, that if the phenomena of reasoning were, finally, in their basic traits, parallel to those of moral conduct, it was, because "reasoning is, essentially, just as moral conduct, a thought submitted to self-control", but he really tried to analyze, in a very detailed way (e.g. CP 1.606), where such a normativity of our logical inferences could come from, using neither a Platonist-Fregean standpoint, nor a merely psychologistic or naturalistic frameworks, but the tools both of experimental psychology and of his scientific, naturalistic and evolutionary metaphysics – the course of evolution being described as the growth of concrete reasonableness, of the power or efficacy of ideas (CP 1.213), of the *Summum bonum* achieved through an aesthetic contemplation of nature (1.615) – and trying to make sense (very much in the footsteps of Kant's suggestion of the "middle course", i.e. a "*preformation* system of pure reason") of the possible emergence of norms from nature (Tiercelin 1997).

For Peirce then, to claim that logical norms are normative is not to view them either as transcendent facts, in a Fregean sense, or as natural facts, as psychologists think, when they try (cf. Mill or Bain) to reduce the laws of thought to the laws of human psychology and the latter to natural laws. Being norms, logical truths and rules cannot be deduced from or reduced to factual propositions bearing on the nature or constitution of individuals. Rather, they are comparable to the rules of conduct or to moral norms: they are imperatives or prescriptions which we follow. However, some explanation has to be offered for their being so 'irresistible' or self-evident (see CP 3.161). Is Ramsey that far from this, when he notes:

> The three normative sciences: Ethics, Aesthetics and Logic, begin, then, with psychological investigations which lead up, in each case, to a valuation, an attribution of one of the three values: good, beautiful, or rational, predicates which appear not to be definable in terms of any of the concepts used in psychology or positive science. I say "appear" because it is one of the principal problems of philosophy to discover whether this is really the case (whether, that is to say, "good", "beautiful", "rational" (or for that matter

> "true") represent undefinable qualities … It is, of course, possible to take one view in regard of one kind of value and the other view with regard to the other kinds; it could be held, for instance, that whereas goodness and beauty could be defined in terms of our desires and admirations, rationality introduced some new element peculiar to logic, such as indefinable probability relations. But the arguments that can be used are so much the same, that when the alternatives that can be used are clearly stated, any normal mind is likely to make the same choice in all three cases. It would be out of place to discuss goodness and beauty in a book on logic, but it will be one of my chief objects to show that the view, which I take of them, that they are definable in [ordinary factual] natural terms, is also true of rationality and truth: *so that just as ethics and aesthetics are really branches of psychology, so also logic is part, not exactly of psychology, but of natural science in its widest sense, in which it includes psychology and all the problems of the relations between man and his environments*. But this is not a matter which can be settled in advance: logic, tries to discover what inferences are rational; we all have some idea as to what this means, but we cannot analyse it exactly until we have made considerable investigations, which are commonly regarded as belonging to logic which is expected to determine the application but also the analysis of its standard of value (OT, 3–5, *my emphasis*)

Just as Peirce claims that we all have in our minds certain norms, or general patterns of right reasoning, which we can compare, approve or disapprove and take as more or less reliable, Ramsey considers that "the human mind works essentially according to general rules or habits; a process of thought not proceeding according to some rule would simply be a random sequence of ideas; whenever we infer A from B, we do so in virtue of some relation between them. We can therefore state the problem of the ideal as "What habits in a general sense would it be best for the human mind to have?'"(PP, 90). And the answer Ramsey gives is in terms of the "human logic of truth", namely, that "given a habit of a certain form, we can praise or blame it accordingly as the degree of belief it produces is near or far from the actual proportion in which the habit leads to truth" (PP, 92).

## *3. Peirce and Ramsey on the human logic of truth*

For both philosophers, to claim that logical norms belong to norms of rationality, is to claim that they are the rules that must be followed by an ideally rational agent; hence not so much a feature that a system of belief or an agent does in fact have, as a trait that governs our interpretation of a system of rational beliefs and behaviours of individuals (Tiercelin 1997, 45). They are inferential norms, governing what we can expect an agent to believe, if he has certain beliefs (e.g., that he has no contradictory beliefs). However, they are also norms that are due to the very truth of the agent's beliefs – otherwise, one could not understand that they should function as norms, namely, that they seem to have some kind of necessity (and self-evidence). This close link between belief and truth is expressed by Peirce when he stresses that it is somewhat redundant to say p is true and to believe that p. Indeed, why should we speak of the notion of truth at all? To tell an inquirer: believe only the truth, is useless. When one has reached a stable belief and put an end to the irritation of doubt, it is purely tautological to say that the believer has reached truth (CP 5.416). When one asks what is the meaning of "true", Ramsey also contends, "it seems to me that the answer is really perfectly obvious, that anyone can see what it is and that difficulty only arise when we try to say what it is, because it is something which ordinary language is rather ill-adapted to express." (OT, 9). As Sahlin has shown, we find in "Truth and Probability" (1926), one of Ramsey's first essays "really imbued with the basic ideas of pragmatism"(1990, 3), together with "Facts and Propositions" (1927), and the note "Knowledge" (1929); all the ingredients composing Ramsey's attitude towards truth, belief, knowledge and probability.

Both Peirce and Ramsey take it that truth must be submitted to an examination of its meaning, to be given a real definition, not a nominal or abstract one, to clear up what Ramsey called the "linguistic muddle" surrounding the concept of truth (PP, 39). "What is truth?" is elusive because it is the wrong question. It is indeed obvious that "It

is true that Caesar was murdered" means no more than that Caesar was murdered. The only real question is what a belief – for example, the belief that Caesar was murdered – is (Sahlin 1990, 3). So; "True" is just a redundant and unnecessary word. In many respects, Peirce and Ramsey join the so-called "deflationist" theories of truth (On Peirce, see Misak 1991, 1998, Tiercelin 1993a, 106ff, 2005, Hookway, 2000, chaps. 2, 3 and 4. On Ramsey, see Dokic & Engel 2001, 31–37, 2002, 18–26).

However, it also means (as shown by the example of the chicken's belief of a poisonous caterpillar) that, "if we have analyzed belief, we have solved the problem of truth" (PP, 39). Hence, we should define the meaning of a belief (i.e., in Peircian terms, a causal disposition to act) "by reference to the actions to which asserting it would lead", that is "by reference to its possible causes and effects" (PP, 51), and it is here that the "pragmatist" twist comes into the picture: if truth does not "need" a definition, it is because it is in a way "metaphysically neutral", and can make sense only so far as "the notion of truth is bound up with the notions of assertion and belief", a point which is underlined by all pragmatists (Peirce, James, Dewey; see Tiercelin 2005, 2014d, for a detailed analysis, and which is glaring in Peirce's approach, as stressed by Misak 2004, 7). Importantly, this shows many things:

*First*, that one should take care of the coherence of our beliefs, for our statements are true because they belong to an integrated ("harmonized") system of beliefs. How can we deny, Ramsey contends, that the truths we seek "[cannot consist in a relation between them and something outside but] must lie inside the system of [our] beliefs, not in a relation between them and an unknowable reality" (OT, 39–40)?

However, one should also, *in the second place*, resist a straightforward "coherentist" temptation, for it creates an "entirely illusory difficulty":

> When we decide that the earth is round, we are not judging that our belief corresponds to a fact but rather that it is the only one which is coherent with our beliefs about ship's disappearing below the horizon, etc. The truth which we

> suppose our belief to have must therefore consist not in correspondence with fact but in coherence with other beliefs. The mistake lies, as we have seen, in supposing that in judging that the earth is round we are thinking about beliefs at all; neither our final judgment that the earth is round nor our initial beliefs that ships disappear, etc. are objects of our thought at all. We are not thinking about our own thinking but about ships disappearing and the earth being round; if coherence comes into our thought at all it is the coherence of reality not the coherence of our beliefs. (OT, 40)

> Suppose the question arises, whether two regiments of soldiers which I have seen or read about have uniforms of the same colour, and I think "The uniforms of the Northshires are red and the uniforms of the Southshires are red, so they have the same colour". Then the relation which I affirm in so concluding is one between the actual colours or uniforms, not one between my opinions about them. My conclusion is founded no doubt on my opinion as to what each colour is, but [is not about those opinions but] states a relation which I believe to hold not between my opinions (or at least not merely between them) but between the real colours. (OT, 38–39)

Indeed, "when we say that other people's opinions are true or false, we mean that they do or do not correspond to the facts" (OT, 39). In other words, our beliefs could well be coherent but false: "The beliefs of a man suffering from persecution mania may rival in coherence those of many sane men but that does not make them true" (OT, 94). This is indeed the reason why even if "evidently we require of our coherent system that it should embody as many as possible of the things we instinctively believe", as Ramsey notes, "we should not regard as plausible a historical system, however coherent, which contradicted all our memories and held that memory was an illusory faculty or even that it went by contraries." (OT, 64). This being said, it does not mean that correspondence should be favoured instead of mere coherence, for the truisms about correspondence are empty, superfluous, providing no interesting information on the pragmatic meaning of the concept, in having no consequences for our practices. As Peirce says,

one talks of "real character," "object," or "reality" (CP 1.578) where one should instead talk of our assertions, commitments, judgments, beliefs, knowledge, inquiry, and of the role played by induction and probability in our approach to the true. Now, a definition of truth which makes no reference to belief, doubt or inquiry is just empty. It is a mere "nominal definition," useful only to those who have never encountered the notion of truth (Misak 1991, 38). If truth is the aim of inquiry, then the correspondence theory leaves inquirers completely in the dark as to how they should conduct their investigations. The aim is not "readily comprehensible" (CP 5.578).

All the same, and *in the third place*, if to a certain extent, correspondence is but a truism, a "platitude" (OT, 12; Hookway 2000, 82), such a "deflationism" about truth is a criticism less of the idea of correspondence than of a metaphysical realism which entertains the illusion of a possible agreement with a real, totally independent of, or transcendent to what we might know of it, a reaction which is central both in Peirce's own "Scotistic" and "scholastic" version of realism (CP, 6.231, 1.27n1) and, later on, of Putnam's primary "internal" realism (RHT, 49) (Tiercelin 1986, 1993a, 11ff, 1993b, 56ff, 2004d). And, as I have shown elsewhere, fruitful links may be drawn between Peirce and Ramsey, when it comes to their respective emphasis on some features which are explicitly viewed by Peirce as part and parcel of a meaningful and straightforward «realistic» attitude, convinced as he was that "pragmaticism could hardly have entered a head that was not already convinced that there are real generals" (CP 5.503): in particular, the importance of generality or the usefulness of thinking in general terms (compare Peirce CP 4. 530, 5.312, 5.425) and Ramsey (PP, 236, OT, 30, OT, 95–96); the recognition of some irreducible indeterminacy and vagueness (compare Peirce CP 4.344, 5.453, 6.348) and Ramsey (PP, 6–7; see Tiercelin 2004d).

For all these reasons, Ramsey's "pragmatist" theory of truth is rather different from the redundancy theory of truth credited to him: in particular, as Sahlin has observed, although the chicken problem can be seen as a decision problem of maximizing one's subjective expected

utility (a choice between the two actions: (i) eat the caterpillar; (ii) refrain from eating the caterpillar) so that we can "use Ramsey's theories of subjective probability, utility and decision to solve it", it also shows that "a truth problem is not one of degrees of belief, but of full belief." "We want to make clear what is meant by saying that the chicken believes fully, i.e. believes that the caterpillar is poisonous. What it means is that the chicken refrains from eating the caterpillar: an action that is useful if and only if the caterpillar is poisonous (and the chicken wants to avoid an upset stomach". (ibid.) From which Sahlin rightly concludes that "this is the gist of Ramsey's theory of truth. It is an obvious example of a pragmatist theory of truth, but also a type of rule-following epistemology. Having a true belief is having a more or less complicated rule, which, if put to use, always leads to success" (Sahlin 1997, 67). So, even if in "Truth and Probability" Ramsey "laid the foundations of the modern theory of subjective probability" and "showed how people's beliefs and desires can be measured by use of a traditional betting method" (Sahlin 1997, 66), it is also "especially fruitful to look upon [that paper] as a theory of rule-following", telling us that "we can describe a person's actions in terms of rule-following "(ibid.), and to connect Ramsey's views on probability to his views on truth, as presented in Facts and propositions."

Now, such a view is reinforced by Ramsey's view on knowledge, which is not merely equated with true, justified (or certain) beliefs, but with reliable ones. As is spelled out in "Reasonable degree of belief" (1928) and "Knowledge" (1929, 110), "a belief, being a map by which we steer, being a rule to follow, must guide our future actions. A full belief, obtained by a reliable method, is definitely not knowledge if it leads us on the wrong track; to be knowledge it must help us to avoid errors. Thus knowledge is simply not true justified belief but rather "a special type of rule-following activity" (Sahlin 1997, 68), and Ramsey defines the notion of reasonable degree of belief by following Peirce's definition: belief is a habit, not an individual judgment produced on a specific occasion, and it is reasonable when the proportion of cases in which this

habit leads to truth is high (PP, 97). Hence, a "causal" – also "functionalist"(see Dokic & Engel 2002, 25) – and "reliabilist" definition of knowledge: a belief is knowledge if it is obtained by a reliable process and if it always leads to success (Sahlin 1990, 3; 1997, 68; 1991, 132–49), success itself being quite different from a mere equation of truth with utility, as should be clear, for example, from Ramsey's harsh criticism of W. James for whom "truth is the expedient in the way of our thinking" and who includes in the truth conditions of the belief in hell the set of all its consequences direct and indirect, making the truth of that belief depend not "on the question whether hell really exists, but on something quite different" (OT, 91–92). Now, "a belief is not true because it is useful, but useful because it is true. Our beliefs are true when our actions are successful, and vice versa, but neither can be reduced to the other. A belief is a successful disposition to act, if and only if the belief is true. In other words: p is true iff p is useful and p is useful iff p is true."(Dokic & Engel 2002, 43–44). Again, and in keeping with the correspondantist elements Ramsey wants to preserve in his account of truth (OT, 11), beliefs have the function of representing real states of the world; they are "maps by which we steer"(PP, 146). They would not be maps if they did not have the function of describing the environment, and they would not be "rules for action", if they did not impose on knowledge the condition of success, implying some "self-control", namely, "not acting on the temporarily uppermost desire, but stopping to think it out", and "forming as a result of a decision an habit of acting [...] in a definite way adjusted to permanent desire". Like Peirce, Ramsey insists, on the one hand on linking truth to reality and, on the other hand, on such second-order general habits of forming general habits about our reliable ways, through which we are justified, in the same way as we are inductively justified to believe our inductive beliefs (Dokic & Engel, 2002, 29–30) and to take induction as the right "human logic of truth":

> We are all convinced by inductive arguments, and our conviction is reasonable because the world is so constituted

> that inductive arguments lead on the whole to true opinions. We are not, therefore, able to help trusting induction, nor if we could help do we see any reason why we should, because we believe it to be a reliable process. It is true that if anyone has not the habit of induction, we cannot prove to him that he is wrong; but there is nothing peculiar in that. If a man doubts his memory or his perception we cannot prove to him that they are trustworthy; to ask for such a thing to be proved is to cry for the moon, and the same is true of induction. It is one of the ultimate sources of knowledge just as memory is: no one regards it as a scandal to philosophy that there is no proof that the world did not begin two minutes ago and that all our memories are not illusory. (PP, 93)

Thus, if the "logic of consistency" is in keeping with formal logic in the narrow, deductive sense, since it rests upon a criterion of coherence between partial beliefs, the "human logic "tells us how humans should think" and how we can sometimes be "humanly right" to entertain a certain degree of belief on inductive grounds: it should be also an inductive logic, or logic of discovery and at times be prepared to go against formal logic (PP, 87). While acknowledging to be unable to assign to such a view any other meaning than "that reasonable opinion [is to be identified] with the opinion of an ideal person in similar circumstances", Ramsey adds: "What, however, would this ideal person's opinion be? [...] The highest ideal would be always to have a true opinion and be certain of it; but this ideal is more suited to God than to man. We have therefore to consider the human mind and what is the most we can ask for it" (PP, 89–90).

### *4. Chance, love and logic*

Now it becomes clearer why, as far as probability is concerned, and as has been pointed out, such a view of truth and of "human logic" is also part and parcel of Ramsey's denial of the relevance of a theory (like Keynes') of probability which would be based on absolutely a priori probabilities (PP, 86ff):

> If we actually applied this process to a human being, found out, that is to say, on what a priori probabilities, his present opinions could be based, we should obviously find them to be ones determined by natural selection, with a general tendency to give a higher probability to the simpler alternatives. But, as I say, I cannot see what could be meant by asking whether these degrees of belief were logically justified. Obviously the best thing would be to know for certain in advance what was true and what was false, and therefore if any one system of initial beliefs is to receive the philosophers approbation it should be this one. But clearly this would not be accepted by thinkers of the school I am criticizing. Another alternative is to apportion initial probabilities on the purely formal system expounded by Wittgenstein, but as this gives no justification for induction it cannot give us *the human logic* which we are looking for. Let us therefore try to get an idea of a human logic which shall not attempt to be reducible to formal logic. Logic, we may agree, is concerned not with what men actually believe, but what they ought to believe, or what it would be reasonable to believe. What then, we must ask, is meant by saying that it is reasonable for a man to have such and such a degree of belief in a proposition? (PP, 88–89, *my emphasis*)

Although Ramsey remains close to Hume's own "sceptical solution" – "the distinction between good and bad reasoning is that between health and disease" (OT, 123) – we have here a principle of justification of induction, which says that "a type of inference is reasonable or unreasonable according to the relative frequencies with which it leads to truth and falsehood" which is opposed both to Humean scepticism about induction and to Wittgenstein's view, who held in the Tractatus (6.361, 6.363, 6.3631) that one can only give a psychological justification of induction. Again, if such a human logic is not an objectivist frequency conception, but a logic of subjective probabilities – laying the foundations of the foundations of the "Bayesian" or "subjective" conception of probabilities and of decision theory –, anchoring what it is rational to believe into actual rules and habits, Ramsey mentions a way of objectifying them, by associating partial beliefs (which bear on the probability that an individual attributes to an isolated event) to frequencies (which bear on classes of

events) admitting that there are, if not connections between the two kinds of probabilities, at least links between the degree of belief, the objective frequency and the utility of the belief. Unlike the other main theorist of the subjective approach, de Finetti, Ramsey does not hold that the meaning of the word "probable" is purely subjective. In other words, "his subjectivism is not incompatible with the claim that beliefs are real dispositions, correlated with objective facts." (Dokic & Engel 2002, 12–13).

Thus we can see why Ramsey's views on probability and chance are not, contrary to one's first reaction, in utter opposition with Peirce'own empirical, anti-subjective conception of chance, in terms of frequencies (CP 2.747) and, later on, of propensities (CP 8.225, 2.664, 8.380), and why, also, the reader of the volume *Chance, Love and Logic* may have felt close, in many ways, to the logician of Milford on those three related issues which take us back again to some links Ramsey – just as Peirce – draw between logic and ethics, and to the view that, in the end: "It is impossible to be completely and rationally logical except on an ethical basis" (CP 2.198) (Tiercelin 2005, 146ff).

Of course, Ramsey's instrumentalist reading of causal laws is well known (PP, 160–161), together with his instrumentalist (or even fictionalist) view that theoretical statements in science are hypotheses which are neither true nor false, but rules or axioms from which one can derive observational consequences (Sahlin 1990, 146–151). Again, Ramsey's anti-realistic reading of counterfactual conditionals, and (at first sight at least) of universals, makes it hard to think that, had he known them, he would have espoused Peirce's realistic views on both counts: namely – and this is the heart of his "Scotistic" conception of universals and of real possibilities which have a form of esse in futuro – 1. that there are real universals, among which habits of nature, "would be's" or dispositions; so the hardness of a diamond is a real fact, because the diamond would resist to pressure if it were hit. And 2. that there are real genuine laws, and that counterfactual conditionals express facts. However, as I have shown elsewhere (and in keeping with some suggestions already made by Dokic & Engel 2002, 42), Ramsey's position in

"Universals" (1925) may itself receive, up to a point, a realistic reading. And most of all, we should not underestimate the ongoing and explicit invocations by Ramsey, of the success of our actions, or, as we saw, of our – if not "formally logical", at least "reasonable" habits of making inductions (PP, 92–94), nor Peirce's understanding of knowledge as inquiry pursued in the long run. ""We do… believe that the system is uniquely determined and that long enough investigation will lead us all to it. This is Peirce's notion of truth as what everyone will believe in the end"; it does not apply to the truthful statement of matters of fact, but to the 'true scientific system'"(PP, 161). Hence, even if causal laws do not describe universal facts, they nevertheless describe regularities upon which our theories will (or would), in the end, converge. Thus it is not surprising that Ramsey "considered himself a disciple of Peirce on many points", "in particular on the fact that it is possible to establish a logic of coherence for probable inference as well as it is for deductive inference"(Levi 1997a, 46).

Of particular interest for us here, are the ethical consequences of such views on probability such as may be found in "The Doctrine of Chances", a paper praised by Putnam for Peirce's acute perception of the depth of the problem of objectivity in ethics and for seeing, better than anyone, that ethical justifications cannot be understood in a purely instrumental way, precisely because they rest on certain norms of rationality (MFR, 80ff; WL, 160–161). Peirce explains why altruism is called by our logic, and why the practical choice made by a person confronted to the dilemma of choosing, in a package, the card that would bring him eternal felicity, in one case, and everlasting woe, in the other case, cannot be performed on a mere utilitarian basis (W 3, 282). Even if, in probabilistic (frequentist) terms, we have no reason to choose either solution, in a single case, we do reason in terms of what it would be more reasonable to believe in the long run, and in the interest of the community as a whole. What guides us in our choice then, is, to a certain extent, indeed, the utilitarian norm (also present in contemporary theories of decision or rational choice): always act so as to maximize the esti-

mated utility (Putnam WL, 161); but no appeal to that rule could be understood in case one did not presuppose that what a rational person pursues in any action is not his own benefit, but what might benefit mankind (or the community of rational investigators) in the infinitely long run. To act otherwise would mean to be "illogical in all one's inferences". For Peirce, "one can only be rational if one identifies himself psychologically with a whole ongoing – in fact a potentially infinite – community of investigators." (Putnam MFR, 83, Tiercelin 2002a).

Putnam is wrong to identify Peirce's solution to the fact that we would recognize the constraining strength of norms which do not possess a satisfactory instrumental justification in terms of our own purposes through a sort of "normative reflection on our practice" (WL, 168), when, as a matter of fact, altruism is less the result of any rational calculus or justification than it appears rather as "immediately" rational. Rather, for Peirce, "that we trust our logical sentiments can be a sign of our wisdom and rationality; our instinctive sense of which actions and reasonings are to be trusted can reflect our grasp of what is required of a reasonable agent, the awareness that our sentimental attunement to the demands of reason exceeds our intellectual understanding of what rationality involves." (Hookway 2000, 239, Tiercelin, 2005, 172ff). Hence, if there is indeed for Peirce a "primitive conception of rationality", it is in the sense in which the "social principle is rooted in our logic", and in which such a sentiment is imperatively required by logic, something which has little to do with any prescription, but works rather (as it does also, with James), through a delicate association, involving our whole being, of Sentiment and Rationality which combine in the formation of moral conduct so as to educate not so much a moral sense as a delicate balance between our ideals and our motives, contributing to a "directly felt fitness with things" (James): in particular, unless ethical norms (which are neither "transcendent" prescriptions nor pure cultural products) were, so to speak, "inhabited" by motives, i.e. not so much moved by emotions as shaped by feelings-dispositions, involving evaluations, they could not really lead us to action, function as

genuine regulative principles, i.e. not as mere "hopes", but as "living hopes" (CP 7.506) (Peirce's equivalent to James's "living options"). And our moral ideals would then rather look like the categorical imperative of the "transcendental apothecaries", not much more valuable than the "barking of a cur" or "the hooting of an owl"(CP 5.133) (Tiercelin 2005, chap.5, 2014d). In other words, "emerging" from nature, norms are never separated from values, our duties being "mere cells of the social organism", and merging into the "universal continuum" (CP 1.673), as is presupposed also by Peirce's evolutionary cosmology: "Every attempt to understand anything— every research supposes, or at least hopes that the very objects of study themselves are subject to a logic more or less identical with that which we employ"(CP 6.189). Thus, if one is rational enough, and scientifically minded, one cannot fail in the long run to discover that reality is such as it is, namely, the growth of some concrete reasonableness which is finally following the action of Love (agapism), the Golden Rule (CP 6.288). This is how the law of habit becomes the law of mind, which again, means that to a certain extent too, the laws of logic may be viewed both as a product of evolution and as the growth of concrete reasonableness (Tiercelin 1997, 40–41).

Now we can understand how "Chance, Love and Logic" can work together. Even if we can never be "sure that the community ever will settle down to an unalterable conclusion upon any given question", we must hope for it, as "the only assumption upon which [we] can act rationally is the hope of success,"(W 2, p. 272). We must have a (regulative though living) hope that for any hypothesis "a prolonged inquiry would declare it to be either true or false" (Misak 1991, 140), otherwise, it would "block the path of inquiry" and make oneself unable to know any positive fact whatsoever (CP 5.603, 5.357, 5.160, 2.650, 2.655, 8.153). The question whether such and such hypothesis is objective may be something that one only knows in the long run. But it is yet possible, now and then, here and there, to have more or less reasonable beliefs on which hypotheses are objective and which are not. (Misak 1991, 141). If such were not the case, then no hypothesis would

have any pragmatic meaning and one would adopt an anti-philosophical attitude. "Though in no possible state of knowledge can any number be great enough to express the relation between the amount of what rests unknown to the amount of the known, yet it is un-philosophical to suppose that with regard to any given question (which has any clear meaning), investigation would not bring forth a solution of it, if it were carried far enough." (W 3, 274).

This is why, on the one hand,

> … all the followers of science are animated by a cheerful hope that the processes of investigation, if only pushed far enough, will give one certain solution to each question to which they apply it … This great hope is embodied in the conception of truth and reality. The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. That is the way I would explain reality. (CP 5.407)

Or:

> I will assume, then, that scientific doubt never gets completely set to rest in regard to any question until, at last, the very truth about that question becomes established … science is foredestined to reach the truth of every problem with as unerring an infallibility as the instincts of animals do their work, this latter result like the former being brought about by some process of which we are as yet unable to give any account. (CP 7.77)

But, on the other hand: "I do not say that it is infallibly true that there is any belief to which a person would come if he were to carry his inquires far enough. I only say that that one is what I call Truth. I cannot infallibly know that there is any truth." (Letter of 1908 to Lady Welby, in Peirce 1958, 398). Or: "We are therefore bound to hope that, although the possible explanations of our facts may be strictly innumerable, yet our mind will be able, in some finite number of guesses, to guess the sole true explanation of them. That we are bound to assume, independently of any evidence that it is true." (CP 7.219).

Not only does Peirce not claim that one will reach

agreement on all questions: opinion can oscillate throughout generations before one comes to a permanent fixation: "The perversity or ignorance of mankind may make this thing or that to be held for true, for any number of generations, but it cannot affect what would be the result of sufficient experience and reasoning. And this is what is meant by the final settled opinion" (W 3, 79). He also observes that it is logically possible that inquiry should go on indefinitely without producing a final answer, but that only shows that to some questions there is no answer, not that for all question which is capable of receiving one, one will not reach an ultimate answer. In other words, one can never demonstrate the exclusion, on any topic whatever, of the logical possibility of error: inversely, the fact that it is logically possible for a thesis to be false, does not imply that it should be doubted. It is always possible too that one should come to an agreement on the false and not on the true, but this is very unlikely. And, obviously, contrary to a presupposition rendered necessary by the success science encounters in the settlements of conflicting opinions:

> If we think that some questions are never going to get settled, we ought to admit that our conception of nature as absolutely real is only partially correct. Still, we shall have to be governed by it practically; because there is nothing to distinguish the unanswerable questions from the answerable ones, so that investigation will have to proceed as if all were answerable. In ordinary life, no matter how much we believe in questions ultimately getting answered, we shall always put aside an innumerable throng of them as beyond our powers. We shall not in our day seek to know whether the centre of the sun is distant from that of the earth by an odd or an even number of miles on the average; we shall act as if neither man nor God could ever ascertain it. There is, however, an economy of thought, in assuming that it is an answerable question. From this practical and economical point of view, it really makes no difference whether or not all questions are actually answered, by man or by God, so long as we are satisfied that investigation has a universal tendency toward the settlement of opinion. (CP 8.43)

All in all, the only definition or real meaning of truth lies in its capacity to determine, in the context of inquiry,

which, among our beliefs, resist doubt and are stable (CP 5.416, 5.375), and Peirce's conception is such that he who searches it may be able and forced to adopt it. The human logic of truth he defends goes hand in hand with the view that "Real pragmatic truth is truth as can and ought to be used as a guide for conduct." (Ms 684, 11, quoted by C. Misak 1991, 159). Such a conception, as Misak observes, presents at least three advantages: "to provide the rational framework for inquiry to proceed" (hence, I would say, it is genuinely "logical"), to "make sense of the practice of inquiry as the search for truth", as something which is not transcendent, beyond inquiry, but accessible (hence, it is, I would add, genuinely "human"), and finally "to justify a methodology" by encouraging the inquirer to put his beliefs to the test of experience. "Peirce's distinctive contribution to debates about truth is to see that, if the aim of inquiry is to get true beliefs, then truth must be thought of as the best that inquiry would do, given as much time and evidence as it takes to reach beliefs which would not be overturned" (Misak 1991, 154).

So it is important to bear in mind that Peirce's position does not require that we can ever reach, or even make sense of, a state of perfect evidence. It requires only "that we can reach a state where no further evidence would disturb the belief that we have arrived at." (Hookway 2000, 49). Which, in Peirce's mind, means that, even if we have reached a point at which our opinions as responsible inquirers are not going to be disturbed by further inquiry, "we never have any absolute guarantee that this position has been reached." (ibid.) No matter how confident we are that we have the truth, further experience could surprise us and oblige us to throw all our beliefs overboard. So truth is "connected to human inquiry (it is the best that inquiry could do), but it goes beyond any particular inquiry (it is not simply the upshot of our best attempts)" (Misak 1998, 408). And this is why what we call in science "established truths" are simply "propositions into which the economy of endeavor prescribes that, for the time being, further inquiry shall cease."(CP 5.589). Relying on his propensity account of probabilities, Peirce compares this rather to the throwing of a pair of dice: when you throw a

pair of dice, you can be sure that it will not fail to obey at one moment its would-be, its propensity to fall on a double six, although this does not imply any logical necessity. (CP, 4.547n1; 7.35) (Levi, 1980a, 1997a, 40).

Thus Peirce is neither an unbound idealist nor a fanatical anti-realist: rather, he constantly oscillates between a more optimistic attitude and an attitude which, in many respects, due to the radical epistemic and ontological fallibilism he also defends, only separates him, at times, from skepticism by a hair's breadth (Hookway 1985, 73), even if such scepticism is softened by a kind of conservative sentimentalism and a critical commonsensism with Reidain and Kantian accents. Now, to a very large extent, this is also the kind of attitude we find in Ramsey, although the king of sceptic he comes closer to, resembles more a Carneades, in the view that probability is indeed the guide in life (See in particular Ramsey, OT, 58, 63). "By knowledge, Ramsey notes,

> … we mean justified confidence, but we regard ourselves as justified in placing complete confidence in arguments that do not begin to amount to strict demonstration. It is even clear that arguments of the sort usually accepted as a sufficient condition for knowledge not only may but sometimes actually do lead to erroneous conclusions. For what scientific proposition is better established than that a man once dead cannot come to life again? And for what fact of ancient history is there better evidence than for the Resurrection of Our Lord?" (OT, 58)

And yet, "one or other of these strong lines of argument must in this instance lead to a false conclusion". This is why, Ramsey contends: "the 'contest of opposite improbabilities', whichever way we resolve it, shows that it is on improbabilities and not on impossibilities that our knowledge is founded (my emphasis). The truth is that we accept as giving knowledge any argument of sufficiently high probability: a confident judgement based on such an argument from known premises is regarded as knowledge when, as is usually the case, it is true." (OT, 58). Just as Ramsey insisted on the need for a human logic to assist the logic of consistency, he notes that we should be careful not

to draw unreasonable or "entirely mistaken" conclusions from the unquestionable fact that "direct knowledge" is "obviously something we can never achieve except perhaps in the simplest cases", such as: "If there is no apprehension, no infallible mode of knowledge, … , what right we can have to be certain of anything?" (OT, 59). Indeed, "in the first place, there is no doubt that people do make mistakes as to the conclusiveness of evidence: and not merely stupid people but even the cleverest, as any teacher or mathematical analysis shows. This fact, with the risk of error which it involves, must simply be faced; it cannot be denied … Our fallibility cannot, therefore, be explained away". However we should not take its consequences as "really so disastrous":

> I see clearly before me a book-case; does the fact that men are occasionally the victims of illusions mean that I have no right to be certain that what I see is indeed a bookcase? Does the fact that men sometimes make mistakes in addition mean that if I and an opponent have arrived independently at the same totals for one bridge scores, we may not be certain that we are right? *To such questions common sense gives us a perfectly clear answer; illusions are so infrequent that it is far best for men to be certain that their judgments of perception are true, and to act accordingly*. The possibility of illusion might be allowed for by their having not a complete conviction but a conviction just a minute fraction, say one part in a million, short of certainty; and in a sense that is what they all have. I say, for instance, that I am certain that what I see is a bookcase, but I am not so absolutely certain that I could not conceivably be persuaded that I was wrong, if anyone could bring forward sufficiently strong arguments to that effect. If one of my friends explained to me that he had been conducting experiments on optical illusions at my expense, I might be convinced by him that what I had seen was not a bookcase at all. If, however, I put an absolute trust in my judgements of perception, I could only come to the conclusion that my friend was a liar, and I should have to take the <b>ally tricks of conjurers for demonstrations of <the>most extraordinary physical phenomena. *So also my arithmetical fallibility means that I should not be wise to put too much trust in an unchecked addition, but if two people get the same answer the chance that they got the same wrong answer by coincidence is negligible and we may*

> *be reasonably certain that they are right. But again not so certain as to refuse to listen to anyone who claimed to have discovered the contrary*. (OT, 59–60, *my emphasis*)

In other words, and very much like Peirce, Ramsey considers that the right attitude here consists in making "such judgements with practical certainty which is not however so complete, that we might not be brought to abandon them if they came into conflict with other beliefs" (OT, 62–63).

## *Conclusion: Peirce and Ramsey on the right conduct of life*

Obviously, Ramsey and Peirce have a lot in common, not only in their views on truth, belief and knowledge, but also when it comes to their conceptions of Logic, Ethics, rational normativity, and even, chance. And to a large extent, this draws Ramsey more on the "realistic" side of Peirce than it draws Peirce to the more Humean, instrumentalistic, and projectivist, "quasi-realistic"" side which is often said to characterize Ramsey (Blackburn 1980, 1993, 75, 1998). I do not deny that such anti-realistic elements are present in several Ramseyan doctrines: to believe that p is to be disposed to act as if p were true; our probability judgements are projections of our degrees of belief, and our theories are but predictive instruments. Unlike atomic beliefs, general beliefs such as "All men are mortal" are neither true nor false, they lack any cognitive content, even if Ramsey's "pragmatist" twist makes him think that this does not imply that they are meaningless (Sahlin 1997, 72), and (in a spirit analogous to Wittgenstein) they do not express maps but merely rules, or instructions for action or attitudes that we take towards the propositions, rules for the formation of judgements, or rather "variable hypotheticals" which we cannot negate, but only disagree with (PP, 148–9) (Sahlin 1995). Again, I am not denying that there are "expressivist" elements when it comes to the characterization of Ramsey's views on ethics, and on the "Hume-Ramsey" theory of rationality, provided one takes Ramsey as interpreting the

Humean element (i.e. instrumental rationality is all rationality), in a non reductive way, and as a kind of formal coherence and ordering among preferences and actions (Gibbard 1990, 10–18, Dokic & Engel 2002, 18). As we have seen too, Ramsey seems at times sceptical about Philosophy or Ethics (PP, 247). Now "Philosophy must be of some use, and we must take it seriously; it must clear our thoughts and so our actions. Or else it is a disposition we have to check, and an inquiry to see that this is so; the chief proposition of philosophy is that philosophy is nonsense. And again we must take seriously that it is nonsense, and not pretend, as Wittgenstein does, that it is important nonsense." (PP, 1) Or also: "But what we can't say, we can't say, and we can't whistle it either." (PP, 146).

However, I would like to suggest that most of such elements or passages are the unavoidable results of the basically fallibilist while at the same time anti sceptical standpoint which Ramsey, together with all pragmatist (Peirce, James, Dewey and Wittgenstein included) adopt, which Putnam perfectly identified as being "the unique insight of American pragmatism" (WL, 152), but which makes it so hard also for them to find a middle way, without being stuck "between the rocks of fallibilism and the whirlwinds of scepticism" which "both sound insane", as D. Lewis noted. As I have shown elsewhere, this may explain why, in the various parries the pragmatists propose to the sceptical challenge, we still find many sceptical components too. Thus, Wittgenstein's criticism of the Cartesian scenario and his attack against radical doubt does not so much comfort a straightforward "realist" or "pragmatist" reading than a basically neo-Pyrrhonian attitude: see his diagnosis of the situation (the sceptical illusion is rather "deflated" than "refuted") or the mobile epistemic status he confers to the "hinge propositions" which seems almost impossible to settle in either sense (Fogelin 1994, 219–222, Tiercelin 2005, chap 3, 104ff). For James too, the condemnation of moral scepticism, presented as a sick obsession of the risk of error, is all the more offensive, as the threat of epistemological scepticism and of an inaccessible objective certainty is strong. If James blames the sceptics for risking apraxia, it is also because, if

we are empiricists, we "believe that no bell in us tolls to let us know for certain when truth is in our grasp" and so "it seems a piece of idle fantasticality to preach so solemnly our duty of waiting for the bell." (Will to Believe). James the empiricist is never far from Sextus or Montaigne, and, more generally from scepticism. But, as I have just said, neither is Peirce, the experimentalist, in whom the sentimental conservative is indeed close to the Reidian commonsensist, but in whom also (as in Ramsey) one can find elements resembling more a Carneades, in the view that probability is the guide in life. And for Peirce, the situation is even worse: scepticism has not the peaceful outlook of academic scepticism, since, for him, epistemological fallibilism goes together with a very extreme ontological fallibilism, which brings it closer, at times, to mere dogmatic scepticism. If our knowledge is basically conjectural and provisional, and if it may even be the case that nothing corresponds to our idea of what reality is (CP 4.61), then, we are not far indeed from falling into the sceptic's well. This explains Putnam's emphasis on the narrowness of the way and his own decision no longer to apply fallibilism to every kind of topics: in particular, Putnam contends, there are questions which it seems impossible to revise (such as: Slavery is bad), or some values which seem to have the hardness of "facts" (as: Yeats was a great poet) (EWO, 16). In that sense, even if we are cognitivists, not only should we favour other terms than "justifications", as applied to the ethical domain, but we should also say that justification must end somewhere (CFVD, 131–132).

In such a context, it is all the more interesting to see the differences in the reactions among the pragmatists, compared to a rather similar diagnosis of the situation in terms of our difficulties to avoid scepticism. A kind of neo-Pyrrhonism is the attitude some finally seem to favour (James, Wittgenstein or Putnam): let us "accept" the "manifest image", our Lebenswelt, the world such as we experience it; hence, let us concentrate our best efforts in recovering a form of "natural realism" or "second naïveté", the sense of the ordinary, of the banal which, Putnam holds, "such strange notions as 'objectivity' and 'subjectivity' we have inherited from ontology and epis-

temology have prevented us from doing" (RHF, 270), with the implicit invitation to follow Apelle's recommendation as reported by Sextus: throw the sponge away, burn all our books and prefer to play backgammon and have dinner in a pleasant company, as Hume at times urges us to do.

However pessimistic both Peirce and even more, Ramsey, at times sound, it seems rather clear, from what we have tried to show, that such is not the attitude that any of them would favour, when it comes to the right conduct of Life. For Peirce (who had probably more than Ramsey, strong metaphysical inclinations), if we want to know what reality consists in (and Peirce's inquiry into truth was first motivated by his inquiry into reality), then we should not only aim at truth, but at knowledge. We should aim at explanations and not at mere descriptions. We should be straightforward realists, in ethics, in science, and in metaphysics. And this gives us obligations, in particular, to specify the right forms of inquiry, so as to integrate in one single move deduction, induction and abduction (at times called by Peirce "the logic of pragmatism"), and to make out which one among the several methods would be able to help "scientific intelligence" to reach the truth: namely, the scientific method, which is subject not only to observation and experience – a part which is relatively out of the control of our will – but also to the strict controls and criticism of our normative capacity of reasoning, inference, deliberation, self-control, criticism, criticism of criticism included. And this is why fallibilism should suffer no exception, from Peirce's point of view (which, incidentally, means that one should also be a fallibilist as regards fallibilism).

However, and Peirce and Ramsey here concur, even if knowledge is conjectural and provisional, one can, in fact, should start with probability, elaborate a true "logic of consistency", and the human logic of truth, which has nothing of a second rate logic. Besides, to say that "absolute" assertions are impossible (CP 1.137–1.140), means that this applies both to those who conclude to any ultimate and perfect formulation (CP 1.440) as to those who claim that such a thing cannot be known (CP 1.138) or

that everything is inexplicable (CP 1.139). Thus dogmatic skepticism, too, is an untenable position. Rather, we are right in thinking most often that most scientific opinions are correct (CP 1.9, 6.603). Besides, this is what the economy of research prescribes (CP 5.589, 1.85.). Simply, a proposition may always be refuted and given up overnight CP 1.120), as much because of the evolution of thought as of the evolution of the laws of nature. After all, a theory which could be absolutely demonstrated would not be a scientific theory (CP 5.541)

In other words, from an ethical point of view, it is important to see that fallibilism does not intend to devaluate knowledge and rather protects from dogmatism and dogmatic skepticism, which in the end are but lazy forms of inquiry. Finally it is such laziness or, in a rather similar vein, the irresponsible abstentionist attitude of the pyrrhonian which seem to me utterly foreign both to Ramsey' and to Peirce' versions of pragmatism. Although somewhat Humean and anti realistic, "Ramsey's pragmatism, as Dokic and Engel have rightly emphasized, "is utterly distinct from the vulgar relativistic or deflationist forms of the so-called neo-pragmatism. [...] Pragmatism is not a theory which would dissolve the real into imagery or into the desirable, but it is a merciless criticism of these dissolutions in the name of our real beliefs and desires." (2002, 80–81). In that respect, it might be worth reminding Peirce's recommendation

> We must not begin by talking of pure ideas, – vagabond thoughts that tramp the public roads without any human inhabitation, – but must begin with men and their conversation. (CP 8.112.)

I think we can also get a glimpse of Ramsey's similar humanity and modesty through the joyful description he gives in his paper read to the Apostles in 1925:

> Where I seem to differ from some of my friends is in attaching little importance to physical size. I don't feel the least humble before the vastness of the heavens. The stars may be large, but they cannot think or love; and these are qualities which impress me far more than size does. I take

> no credit for weighing nearly seventeen stone. My picture of the world is drawn in perspective, and not like a model to scale. The foreground is occupied by human beings and the stars are all as small as three penny bits. I don't really believe in astronomy, except as a complicated description of part of the course of human and possibly animal sensation. I apply my perspective not merely to space but also to time. In time the world will cool and everything will die; but that is a long time off still, and its present value at compound discount is almost nothing. Nor is the present less valuable because the future will be blank. Humanity, which fills the foreground of my picture, I find interesting and on the whole admirable. I find, just now at least, the world a pleasant and exciting place. You may find it depressing; I am sorry for you, and you despise me. But I have reason and you have none; you would only have a reason for despising me if your feeling corresponded to the fact in a way mine didn't. But neither can correspond to the fact. The fact is not good or bad; it is just that it thrills me but depresses you. On the other hand, I pity you with reason, because it is pleasanter to be thrilled than to be depressed, and not merely pleasanter but better for all one's activities. (Epilogue, PP, 249–250)

Just in passing, Nils-Eric Sahlin noted that "Ramsey's mother was active in politics and he inherited from her a very profound social awareness" (Sahlin 1990, 222). This is indeed the impression we get from reading him. As Sahlin insisted on, too, it makes perfect sense to claim, that "Ramsey must have influenced Wittgenstein", rather than the other way round, and that, at any rate, Ramsey used this theory of probability and truth in order to "solve" problems, not to "dissolve them" (Sahlin 1997, 69). Last but not least, the "pragmatist" twist Ramsey gave to his works, at one point, may also explain why, at about the same time as he was paying more and more attention to Peirce, he finally came to think that "Wittgenstein had nothing new to offer him" and was "no good for (his) work"(Sahlin 1997, 64).

## *References*


Blackburn, S. (1980), Opinions and Chances, in D. H. Mellor (ed.).

Blackburn, S. (1993), *Essays in Quasi-Realism*, Oxford, Oxford University Press.

Blackburn, S. (1998), Wittgenstein, Wright and Minimalism, *Mind*, 107(425), 157–181.

Blackburn, S. (2001), *Being Good: a short introduction to ethics*, Oxford, Oxford University Press.

Dokic, J. & Engel, P. (2001), *Ramsey, vérité et succès*, Paris, Presses Universitaires de France.

Dokic, J. & Engel, P. (2002), *Ramsey, Truth and Success,* London, Routledge.

Fogelin, R. (1994), *Pyrrhonian Reflections on Knowledge and Justification*, Oxford, Oxford University Press.

Gavalotti, M.–C. (2014), New Prospects for pragmatism: Ramsey's Constructivism, *New Directions in the Philosophy of Science, The Philosophy of Science in a European Perspective Volume 5*, 645–656.

Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Harvard, Harvard University Press.

Gärdenfors, P. & N.E. Sahlin (1982), Unreliable probabilities, risk taking, and decision making, *Synthese* 53, 361–386.

Hookway, C. (1985), *Peirce*, London: Routledge and Kegan Paul.

Hookway, C. (2000), *Truth, Rationality and Pragmatism*:Themes from Peirce, Oxford: Clarendon Press.

Levi, I. (1980a), Induction as Self Correcting According to Peirce, in D.H. Mellor (ed.), *Science, Belief and Behaviour*; Cambridge, Cambridge UP, p. 127–140.

Levi, I. (1980b), *The Enterprise of Knowledge*, Cambridge, Cambridge UP.

Levi, I. (1983), *The Enterprise of Knowledge – An Essay on Knowledge, Credal probability, and Chance*, Cambridge, Mass.: the MIT Press.

Levi, I. (1991), *The Fixation of Belief and Its Undoing*, Cambridge, Cambridge UP

Levi, I. (1995), Induction According to Peirce, in K.L. Ketner (ed.), *Peirce and Contemporary Thought: Philosophical Inquiries*, New York, Fordham University Press.

Levi, I. (1997), Inference and Logic according to Peirce, in J. Brunning & P. Forster (eds.), *The Rule of Reason*, Toronto, University of Toronto Press, 1997, 34–56.

Lewis, D. (1996), Elusive Knowledge, *Australasian Journal of Philosophy* 74, 549–67, reprinted in David Lewis, 1999, 418–445, *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press.

Mellor, D. H. (1980), *Prospects for Pragmatism. Essays in the memory of F. P. Ramsey*, Cambridge, Cambridge University Press.

Misak, C. (1991), *Truth and the End of inquiry*, Oxford, Oxford University Press.
Misak, C. (1998), Deflating Truth: Pragmatism vs minimalism, *The Monist* 81, 407–25.
Misak, C. (2004), Charles Sanders Peirce (1839–1914), in C. Misak (ed.), *The Cambridge Companion to Peirce*, Cambridge, Cambridge University Press, 2004, 1–26.
Misak C., (ed.)(2004), *The Cambridge Companion to Peirce*, Cambridge,Cambridge University Press.
Peirce, C. S. (1923), *Chance, Love and Logic*, M. R.Cohen ed., New York: Barnes & Noble Inc.
Peirce, C. S. (1931–58), *The Collected Papers of C.S. Peirce*, C. Harsthorne, P.Weiss, & A. Burks eds., Cambridge, Mass., Harvard University Press (8 vols.).
Peirce, C. S. (1958), *Selected Writings, Values in a Universe of Chance*, Ph. Wiener ed., New York, Dover.
Peirce, C. S. (1976), *The New Elements of Mathematics* (NEM), C. Eisele (ed.), The Hague, Mouton, (4 vols.).
Peirce, C. S. (1982 –), *Writings of C.S. Peirce: a chronological edition* (W), M. Fisch, C.Kloesel, E.C. Moore eds., Bloomington, Indiana University Press (6 vols. published).
Peirce, C. S. (1992), *Reasoning and the logic of Things* (RLT), K. Ketner ed., Cambridge, Mass., Harvard UniversityPress.
Putnam, H. (1987), *The Many Faces of Realism* (MFR), Open Court, La Salle, Ill.
Putnam, H. (1990), *Realism with a Human Face* (RHF), J. Conant (ed.), Cambridge, Mass.: Harvard. (1994) *Words and Life* (WL), J. Conant (ed.), Cambridge, Mass.: Harvard University Press.
Putnam, H. (2002) *The Collapse of the Fact-Value Dichotomy and Other Essays* (CFVD), Cambridge, Mass., Harvard University Press.
Putnam, H. (2004), *Ethics without Ontology* (EWO) Cambridge, Mass., Harvard University Press.
Ramsey, F. P. (1990), *Philosophical Papers* (PP), DH Mellor ed., Cambridge, Cambridge University Press.
Ramsey, F. P. (1991a), *Notes on Philosophy, Probability and Mathematics* (NPPM), M. C. Gavalotti (ed.), Naples, Bibliopolis.
Ramsey, F. P. (1991b), On Truth (OT), N. Rescher et U Maier (eds.), *Episteme* 16, Kluwer
Sahlin, N.-E. (1990), *The Philosophy of F. P. Ramsey*, Cambridge: Cambridge University Press.
Sahlin, N.-E. (1991), Obtained by a reliable process and always leading to success, *Theoria* 57, 1991, 132–149
Sahlin, N.-E. (1995), On the Philosophical Relations between Ramsey and Wittgenstein, in J. Hintikka & K. Puhl (ed.), *The British Tradition in 20th Century Philosophy. Proceedings of the 17th International Wittgenstein-Symposium*, Vienne, Hölder-Pichler-Tempsky, 1995, 150–163.

Sahlin, N.-E. (1997), "he is no good for my work", On the philosophical relations between Ramsey and Wittgenstein", *Knowledge and Inquiry essays on J. Hintikka's Epistemology and Philosophy of Science*, (M. Sintonen ed.), Poznan Studies in the Philosophy of the sciences and the humanities, 1997, 61–84.

Suppes, P. (2004), Ramsey's Psychological Theory of Belief. *Cambridge and Vienna, Frank P. Ramsey and the Vienna Circle*, A.-M. Gavalotti ed., Vienna Circle Institute Yearbook 12, Springer, 35–54.

Tiercelin, C. (1992), Vagueness and the unity of Peirce's realism, *Transactions of the C. S. Peirce Society*, Winter 1992, vol. XXVIII, n°1, 51–82.

Tiercelin, C. (1993a), *C. S. Peirce et le pragmatisme*, Paris, Presses Universitaires de France, 1993; épuisé; en ligne: http://books.openedition.org/cdf/1985

Tiercelin, C. (1993b), *La pensée-signe: études sur Peirce*, Nîmes, Editions Jacqueline Chambon, 1993; épuisé; en ligne: http://books.openedition.org/cdf/2209.

Tiercelin, C. (1997), Peirce on Norms, Evolution, and Knowledge, *Transactions of the C.S. Peirce Society*, vol. XXXIII, 1997, n°1, 35–5.

Tiercelin, C. (2002a), *Hilary Putnam, l'héritage pragmatiste*, Paris, Presses Universitaires de France, 2002; épuisé; en ligne: http://books.openedition.org/cdf/2010

Tiercelin, C. (2002b), Philosophers and the Moral Life, *Transactions of the C. S. Peirce Society*, vol. XXXVIII,N°1/2, 2002b, 307–326.

Tiercelin, C. (2004a), "Les philosophes et la vie morale", in *L'éthique de la philosophie*, J.-P. Cometti (dir.), Kimé, Paris, 2004a, 15–38.

Tiercelin, C. (2004b) "Peirce, lecteur d'Aristote", in *Aristote au XIXe siècle*, sous la direction de D. Thouard, Lille, Presses Universitaires du Septentrion, 2004b.

Tiercelin, C. (2004c) Abduction and the Semiotics of Perception, *Semiotica*, F. Merrell et J. Queiroz eds, 2004c.

Tiercelin, C. (2004d) Ramsey's pragmatism, *Dialectica* vol. 58 Fasc. 4, 529–547.

Tiercelin, C. (2005), *Le doute en question: parades pragmatistes au défi sceptique*, Paris, Editions de l'éclat, 2005. http://www.amazon.fr/Le-doute-question-pragmatistes-sceptique/dp/2841620948 http://books.google.fr/books/about/Le_doute_en_question.html?id=RPrkSuqZM_QC&redir_esc=y

Tiercelin, C. (2014), *The Pragmatists and the Human Logic of Truth*, in La Philosophie de la connaissance au Collège de France (C. Tiercelin ed.), collection Métaphysique et connaissance, online edition, oct. 2014, http://books.openedition.org/cdf/3652.

Wittgenstein, L. (1969), *Uber Gewissheit* (UG) (On Certainty), G. E. M. Anscombe & G. H. von Wright (eds.), Oxford, Blackwell.

# Epilog

FRANK RAMSEY

Ställd inför uppgiften att behöva skriva en uppsats för Apostlarna fann jag mig som vanligt stå utan något ämne, och jag smickrade mig med att detta inte bara var ett personligt tillkortakommande, utan att det verkligen inte fanns något ämne som var lämpligt att diskutera. Men då jag nyss hade föreläst kring typteorin, slog det mig att i ett sådant påstående skulle ordet ”ämne” behöva begränsas till att bara omfatta ämnen av första ordningen, och kanske skulle det kunna finnas ett möjligt andra ordningens ämne. Och sedan såg jag att det låg färdigt framför mig, nämligen att det inte skulle finnas något ämne att diskutera (av första ordningen).

En allvarlig sak om det är sant. För vad är syftet med Apostlarnas existens, om inte diskussion? Och om det inte finns något att diskutera – men det kan vi ta upp efteråt.

Jag vill inte påstå att det aldrig har funnits något att diskutera, utan bara att det inte finns det numera – att vi verkligen har avgjort allt genom att inse att det inte finns något att veta utöver vetenskapen. Och att de flesta av oss är okunniga om de flesta vetenskaper, så även om vi kan utbyta information kan vi inte på ett givande sätt diskutera dem, eftersom vi bara är nybörjare.

Låt oss ta en titt på de möjliga ämnen som finns att diskutera. De faller, så vitt jag kan se, under rubrikerna vetenskap, filosofi, historia och politik, psykologi och estetik; där har jag, för att inte förutsätta något, skilt psykologin från de andra vetenskaperna.

Vetenskap, historia och politik är inte lämpliga att diskutera, förutom av experter. Andra är helt enkelt i den positionen att de behöver mer information, och innan de

har införskaffat all tillgänglig information, kan de inte göra något annat än ta för givet de uppfattningar som mer kvalificerade personer har. Så finns det filosofi, men även filosofin har blivit för teknisk för lekmannen. Förutom denna nackdel, så har den största moderna filosofen kommit fram till slutsatsen att det inte finns något sådant ämne som filosofi: att det är en aktivitet, inte en doktrin, och att den, istället för att besvara frågor, bara syftar till att bota huvudvärk. Man skulle kunna tro att det utöver denna tekniska filosofi med logik i centrum fanns ett slags populärfilosofi som handlade om sådant som människans relation till naturen och moralens mening. Men varje försök att behandla sådana ämnen på allvar reducerar dem till frågor inom antingen vetenskap eller teknisk filosofi, eller leder till att de mer omedelbart inses vara nonsens.

Tag till exempel Russells nyligen utgivna föreläsning ”Vad jag tror” (”What I Believe”, 1925). Han delade upp filosofin i två delar: naturfilosofin och värdefilosofin. Hans naturfilosofi bestod främst av slutsatser inom den moderna fysiken, fysiologin och astronomin, med en lätt inblandning av hans egen teori om materiella föremål som ett särskilt slags logisk konstruktion. Dess innehåll kunde därför diskuteras endast av någon med en adekvat kännedom om relativitetsteorin, atomfysik, fysiologi och matematisk logik. Den enda återstående möjligheten till diskussion i anslutning till denna del av hans uppsats skulle vara den betoning han lade på vissa saker, till exempel skillnaden i storlek mellan stjärnorna och människorna. Jag ska återkomma till den saken.

Hans värdefilosofi bestod i att säga att de enda frågorna om värde rörde vad människor önskade sig, och om hur deras önskningar skulle uppfyllas, och så fortsatte han med att besvara dessa frågor. Därmed blev hela ämnet en del av psykologin, och diskussionen av det skulle vara en psykologisk diskussion.

Naturligtvis kunde man ifrågasätta hans centrala påstående om värde, men de flesta av oss skulle hålla med om att det godas objektivitet var något som vi hade avgjort och avfört från diskussionen med den om Guds existens. Teologi och Absolut Etik är två berömda ämnen som vi nu förstår inte handlar om något reellt.

Etiken har då reducerats till psykologin, och det tar mig då till psykologin som ett ämne för diskussion. De flesta av våra möten skulle kunna sägas handla om psykologiska frågor. Det är ett ämne vi av praktiska skäl alla är mer eller mindre intresserade av. När vi behandlar det måste vi skilja ut den egentliga psykologin, som är studiet av mentala händelser med syftet att slå fast vetenskapliga generaliseringar, från att bara jämföra våra egna upplevelser utifrån ett personligt intresse. Det avgörande provet är om vi skulle vara lika intresserade av denna upplevelse ifall det var en främlings som vi är om det rör sig om vår väns – om vi är intresserade av det som vetenskapligt stoff, eller bara utifrån personlig nyfikenhet.

Jag tror att vi sällan, om någonsin, diskuterar fundamentala psykologiska frågor, utan att vi mycket oftare helt enkelt jämför våra olika upplevelser, vilket inte är ett sätt att diskutera. Jag tror att vi inte inser tillräckligt hur ofta våra argument är av formen: A: ”Jag åkte till Grantchester imorse.” B: ”Nej, det gjorde jag inte.” En annan sak som vi ofta gör är att diskutera vilka slag av människor eller beteende vi beundrar eller skäms för. När vi till exempel diskuterar affektkonstans består det i att A säger att han skulle känna sig skyldig om han inte uppvisade en sådan konstans, och att B säger att han inte skulle känna sig ett dugg skyldig. Men detta är inte att diskutera någonting alls, utan bara att jämföra sina anteckningar (även om det är ett angenämt sätt att fördriva tiden).

Genuin psykologi är å andra sidan en vetenskap som de flesta av oss vet alldeles för lite om för att det ska passa sig för oss att ha en uppfattning.

Till sist finns det estetik, inklusive litteratur. Detta hetsar alltid upp oss mycket mer än allt annat, men vi diskuterar det egentligen inte så mycket. Våra argument är så svaga. Vi är fortfarande på nivån ”Den som driver feta oxar måste själv vara fet”, och vi har mycket lite att säga om de psykologiska problem som estetiken egentligen består av, till exempel varför vissa kombinationer av färger ger oss så märkliga känslor. Det vi egentligen tycker om att göra är återigen att jämföra våra upplevelser, en praktik som i detta fall är märkligt lönsam, därför att kritikern kan peka ut saker för andra människor, saker som de, om de är

uppmärksamma på dem, gör att de kan få upplevelser som de värderar vilka de inte annars skulle ha haft. Vi diskuterar inte, och vi kan inte diskutera, om ett konstverk är bättre än ett annat; vi jämför bara de känslor det ger oss.

Jag drar alltså slutsatsen att det inte finns något att diskutera, och denna slutsats motsvarar även en känsla jag har om vanlig konversation. Det är ett ganska nytt fenomen, som har uppkommit ur två orsaker som har verkat gradvis under artonhundratalet. Det ena är vetenskapens framsteg, den andra religionens nedgång; detta har lett till att alla de gamla generella frågorna har blivit antingen tekniska eller löjliga. Denna process i civilisationens utveckling måste vi alla genomgå inom oss själva. Jag själv kom till exempel hit som en förstaårsstudent som njöt av konversationer och argument mer än allt annat i världen, men jag har gradvis kommit att se det som allt mindre viktigt, därför att det aldrig verkar finnas något att tala om utom rena yrkesfrågor och människors privatliv, och inget av dessa är lämpligt för allmänna samtal. Och sedan jag genomgick analys känner jag att folk vet mycket mindre om sig själva än vad de tror, och jag är inte tillnärmelsevis så ivrig att tala om mig själv som jag brukade vara – jag har gjort det tillräckligt för att bli uttråkad av det. Konst och litteratur finns ju fortfarande, men om dessa saker kan man inte argumentera, man kan bara jämföra sina anteckningar, precis som man kan utbyta information om historia eller ekonomi. Men om konsten utbyter man inte information utan känslor.

Detta tar mig tillbaka till Russell och ”Vad jag tror”. Om jag skulle skriva en världsåskådning, skulle jag inte kalla den ”Vad jag tror”, utan ”Vad jag känner”. Detta är förknippat med Wittgensteins uppfattning att filosofin inte ger oss trosföreställningar, utan bara lindrar känslor av intellektuellt obehag. Och, om jag dessutom skulle argumentera mot Russells föreläsning, skulle det inte vara mot det han tror, utan mot de antydningar vi får om vad han kände. Nu kan man ju egentligen inte vara oense med en människas känslor, man kan bara ha andra känslor för egen del, och kanske också se sina egna känslor som mer beundransvärda eller bättre i stånd att främja ett lyckligt liv. Utifrån denna ståndpunkt, att det inte är en fråga om

fakta utan om känslor, ska jag avsluta med några anmärkningar kring saker i allmänhet, eller, som jag hellre skulle uttrycka det, inte saker utan *livet* i allmänhet.

Den punkt där jag verkar skilja mig från vissa av mina vänner är att jag tillskriver fysisk storlek liten betydelse. Jag känner mig inte det minsta ödmjuk inför himlavalvens storlek. Stjärnorna må vara stora, men de kan inte tänka eller älska – och dessa är egenskaper som gör ett vida större intryck på mig än vad storlek gör. Jag begär inga extra förmåner för att jag väger över 100 kg.

Min bild av världen är ritad i perspektiv, och inte som en skalmodell. Förgrunden upptas av människor, och stjärnorna är alla små som trepennymynt. Jag tror egentligen inte på astronomin, förutom som en komplicerad beskrivning av en del av människors och kanske djurs erfarenhet. Mitt perpektiv gäller inte bara rummet, utan också tiden. I sinom tid kommer jorden att svalna och allt kommer att dö, men det är fortfarande långt till dess, och detta faktums diskonterade nuvärde är nära noll. Inte heller är det nuvarande mindre värdefullt för att framtiden kommer att vara tom. Mänskligheten, som upptar förgrunden i min bild, finner jag vara intressant och på det hela taget beundransvärd. Jag finner, åtminstone just nu, att världen är ett tilltalande och intressant ställe. Du kanske finner den deprimerande; jag är ledsen för din skull, och du föraktar mig. Men jag har skäl och du har inga; du skulle ha skäl att förakta mig om dina känslor motsvarade fakta på ett sätt som mina inte gör. Men ingendera kan motsvara faktum. Detta faktum är inte i sig självt gott eller ont; det är bara det att det gör mig upprymd och dig deprimerad. Å andra sidan har jag skäl att tycka synd om dig, därför att det är trevligare att vara upprymd än att vara deprimerad, och inte bara trevligare utan bättre för alla ens aktiviteter.

*28 februari 1925*

Detta föredrag hölls av Ramsey inför en diskussionsförening i Cambridge, The Apostles. Engelska originalet finns i F. P. Ramsey, *Philosophical Papers*, redigerad av D. H. Mellor, Cambridge University Press, Cambridge, 1990.

*Fredrik Stjernberg* (översättning)

IMAGE CREDITS
Cover image: *Mein Kindermädchen*, 1936
© Meret Oppenheim/Bildupphovsrätt 2015
The British Library (p. 157), The British Museum (121), J. Paul Getty Museum (136, 138, 140), The Morgan Library & Museum (137), The National Gallery (143), Wikimedia Commons (114, 116, 118, 123, 125, 128, 134, 142, 145, 147, 149, 154)

OTTO NEURATH
*Gesellschaft und Wirtschaft* (p. 153) can be downloaded at medienphilosophie.net/neurath/Gesellschaft_und_Wirtschaft_1931.pdf

Fri tanke förlag
www.fritanke.se
info@fritanke.se


Design: Johan Laserna
Printed by Media-Tryck, Lund 2015
ISBN 978-91-87935-37-4