Emotion Researcher, 2011, 26(2), 5-7

Modelling Machines That Can Love:
From Bowlby's Attachment Control System to Requirements for Romantic Robots

Dean Petters
University of Birmingham, UK (Computer Science)

Everett Waters
SUNY Stony Brook, US (Psychology)

Aaron Sloman
University of Birmingham, UK (Philosophy and Computer Science)

Most theories of love are drawn in broad strokes. Such theory sketches can be useful but may fail to provide the detail and specificity necessary to formulate and evaluate strong empirical tests. Computational modelling can overcome some of these limitations by being more explicitly and precisely formulated than traditional verbal theories (Wright, Sloman & Beaudoin, 1996, Weisberg 2007). As a consequence, computational models often raise interesting questions about how theoretical constructs might be formulated and inter-related. Implemented in simulations or robots, computational models are also a useful way to explore the completeness and consistency of a theory over time spans and contexts not readily accessible to experiments.

Approaching love as a phenomenon to be modelled, rather than experimentally dissected, challenges us to specify the kinds of information processing subsystems that would be required by any love-capable system (including humans). It also challenges us to consider various architectural arrangements among subsystems (Sloman 2000), how these might change or adapt with experience, and whether the capacity for love might require, as well, appropriate formative experiences.

Aspects of love phenomena particularly salient to a modelling and design perspective include (1) emotional intensity and priority, (2) selectivity and a tendency toward monotropy, (3) continuity combined with only intermittent expression, and (4) priority over and influence on various goal structures. Ethologically inspired work on infant-mother attachment, parenting, and adult-adult relationships highlights a wide range of important proximity seeking and secure base-related behaviours that are hallmarks of our closest relationships (e.g., Bowlby, 1969; Ainsworth, Blehar, Waters, & Wall, 1978; Petters, Waters and Schönbrot, 2010). Such studies provide rich and detailed scenarios that serve as targets for the design process (Petters 2006). They are supplemented by psychometric research highlighting the importance of intimacy, passion, and commitment in adult romantic relationships (e.g., Sternberg 1986).

In addition to building upon clinical and ethological observations, Bowlby (1969) referred to a very broad range of information processing structures and mechanisms to explain the rich set of emotional phenomena present in attachment relationships. In doing so, he borrowed from ethology the concept of behavioural systems, from cybernetics the control systems concept and the idea of internal working models, and from artificial intelligence the notion of algorithms and planning representations.

Working with such constructs, a number of researchers have developed software simulations (e.g., Bischof 1973, Petters 2006) and robots (e.g., Canamero, Blanchard and Nadel 2006) that use one or a few preferred figures as a secure base from which to explore and as a haven of safety when distressed. It would be a significant advance if future models could capture a range of emotional

phenomena wider and more complex than mere distress and comfort and examine their roles as sources of information for, and consequences of, information processing in perceptual and cognitive subsystems. One might expect that this would help clarify phenomena such as grief and mourning processes experienced in response to loss which are beyond the scope of current models.

It is a simple matter for anyone with appropriate programming skills to mimic love-associated phenomena in an arbitrary system. It is much more interesting and informative to design more generalized, cognitively and biologically plausible, subsystems and architectures and examine the extent to which love-like phenomena appear without having been "cooked into" the design. That is, rather than building a "love machine", we seek to build machines that are first of all adaptive, capable of learning and acting in a timely manner in a dynamic world, and secondarily (perhaps as a consequence) are also capable of love.

Sloman (2000) presents a systematic architectural framework, which is intended only as a first approximation to summarising layers of control and cognition produced by evolution. This three layered framework helps explain how love phenomena can result from primary, secondary and tertiary emotions, and also explains how automatic and controlled processes compete for control in emotional episodes. Primary emotions, such as being startled, terrified or delighted, arise as global interrupts in the lower reactive layer of cognitive architectures. Secondary emotions, such as being anxious, apprehensive or relieved, arise from interruptions to deliberative processing in the middle layer. Tertiary emotions arise from disturbances to processing in the higher third layer of cognitive architectures. Examples include loss of attentional control seen in episodes of grief or longing when in intense love. This third 'reflective' level is concerned with managing processes that occur in the lower architectural levels. Loss of control when in love can involve thoughts about the object of love that often intrude upon consciousness when an individual is engaged in other tasks – a breakdown of this process management. So being able to capture the phenomenon that thoughts occur that cannot easily be put out of mind - is not enough. Modelling these management processes comes first - to model how love involves losing control there must be some control to start with. In this framework, all three types of emotion can exist without being observed – giving rise to dormant dispositions. A primary emotion (being startled) may have its normal behavioural response suppressed, e.g. because of anxiety about being detected. Or a tertiary emotion such as intense longing maybe suppressed by other urgent and important concerns, but the disposition to regain control under some circumstances persists. Disturbances or 'perturbations' to the third management level may not involve interruptions in the same sense as lower levels but can involve changes to how effectively processes in other levels can be controlled. Also, not all the perturbations are impairments. Some involve acceleration, redirection, or tighter control that avoids a disaster.

Undoubtedly, some readers will object that computational modelling is necessarily "cold" and unsuited to research on a phenomenon as "warm" and emotional as love. This underestimates the range of conceptual and programming tools available to today's modellers. Moreover, it is more addressed to love mimicry than to love modelling. Whether we can design machines that can love is yet to be determined. We are confident, however, that we can learn a great deal about love by trying to do so.

References

Ainsworth, M., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment*. Hillsdale, NJ: Erlbaum.

Bischof, N. (1975). A systems approach toward the functional connections of attachment and fear. *Child Development, 46*, 801-817.

Bowlby, J. (1969). *Attachment and loss: Volume 1 Attachment*. New York: Basic Books.

Canamero, L., Blanchard, A., and Nadel, J. (2006). Attachment bonds for human-like robots. *International Journal of Humanoid Robotics, 3* (3), 301-320

Petters, D. (2006). Designing agents to understand infants. Ph.D. thesis in Cognitive Science, School of Computer Science, Univ. Birmingham.

Petters, D., Waters, E., & Schönbrodt, F. (2010). Strange carers: Robots as attachment figures and aids to parenting. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 11*(2), 246-252.

Sloman, A. (2000). Architectural requirements for human-like agents both natural and artificial. (What sorts of machines can love?). In Kerstin Dautenhahn (Ed.) *Human Cognition and Social Agent Technology*, in the series "*Advances in Consciousness Research*", Amsterdam: John Benjamins Publishing.

Sternberg, R.J. (1986). A triangular theory of love. *Psychological Review, 93* (2), 119–135.

Weisberg, M. (2007). Who Is a Modeler. *British Journal of the Philosophy of Science. 58*, 207–233.

Wright, I., Sloman A., & Beaudoin L. (1996), Towards a design-based analysis of emotional episodes. *Philosophy Psychiatry and Psychology*, *3*(2), 101-126.